Name: Key Unit 6 - Review (Ch. 23 – 25) Date: ____________

1. **Vacations days.** The distribution of the number of vacation days per year offered by different U.S. companies is skewed to the right.

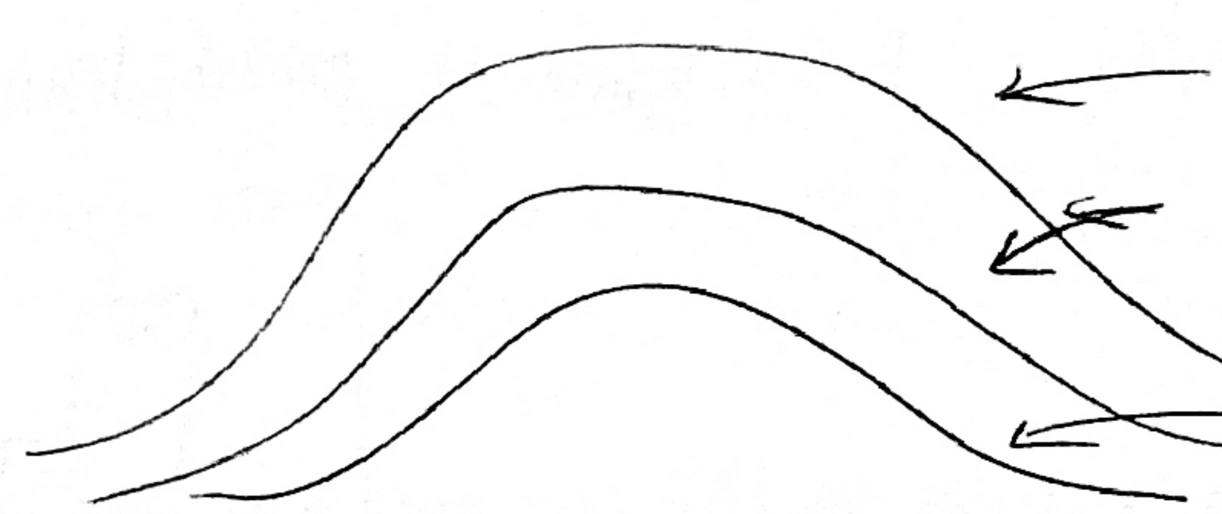
a) We collect data on the number of vacation days from a random sample of 60 companies across the United States. Why is it okay to use these data for inference even though the population is skewed?

we have a large sample size

b) The mean and standard deviation of the 60 companies in our sample were 22 days and 9 days, respectively. Specify the sampling model (shape, center, spread) for the mean number of vacation days of such samples.

t-distribution

$t_{59}\left(22, \frac{9}{\sqrt{60}}\right)$



c) Find a 95% confidence interval for the mean number of vacation days offered by U.S. companies.

Conditions: Random: we have a random sample of U.S. companies

10%: the sample is < 10% of the total # of U.S. companies

Normality: we have a large enough sample size to proceed

→ 1 sample t-interval: $n = 60$; $\bar{x} = 22$; $s = 9$ & $SE(\bar{x}) = \frac{9}{\sqrt{60}} = 1.16$

$\bar{x} \pm t^*_{59} \frac{s}{\sqrt{n}}$ $\quad t^*_{59} = 2.009$

$22 \pm 2.009 \cdot \frac{9}{\sqrt{60}}$

$= 22 \pm 2.33$

$(19.7, 24.3)$

we are 95% confident that the interval 19.7 to 24.3 contains the true mean number of vacation days that are given U.S. companies.

d) Explain what "95% confidence" means in this context.

If many random samples of size 60 were taken, 95% of the confidence intervals produced would contain the actual mean number of vacation days offered by U.S. Companies.

A)

2. Too much TV? A father is concerned that his teenage son is watching too much television each day, since his son watches an average of 2 hours per day. His son says that his TV habits are no different than those of his friends. Since this father has taken a stats class, he knows that he can actually test to see whether or not his son is watching more TV than his peers. The father collects a random sample of television watching times from boys at his son's high school and gets the following data:

1.9 2.3 2.2 1.9 1.6 2.6 1.4 2.0 2.0 2.2

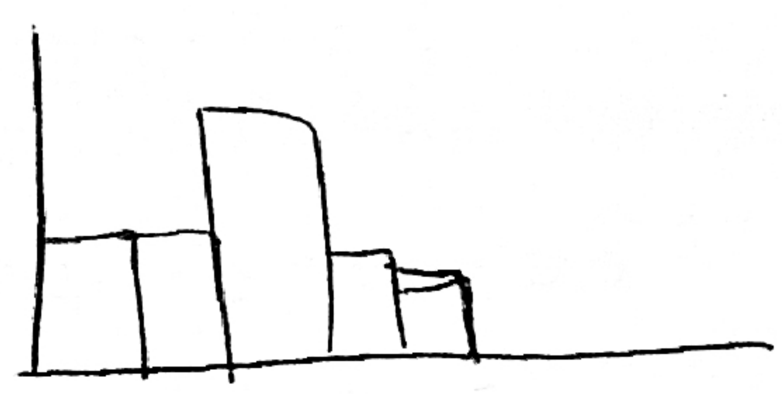
Is the father right? That is, is there evidence that other boys average less than 2 hours of television per day? Conduct a hypothesis test, making sure to state your conclusions in the context of the problem.

① $H_0: \mu = 2.0$ hours
$H_a: \mu < 2.0$ hours

② Random: Boys from the high school were randomly selected.

10%: The sample is less than 10% of all boys at the high school.

Normality: A histogram of credit hours is unimodal and roughly symmetric



→ conditions met → 1-sample t-Test

③ $df = 10-1 = 9$

$\bar{x} = 2.01$

$S_{\bar{x}} = 0.345$

$$t = \frac{2.01-2}{\frac{0.345}{\sqrt{10}}} = 0.0917$$

$$P(t < 0.0917) = 0.536$$

④ Since p-value is high, we fail to reject Ho. There is insufficient evidence the son is watching more T.V., one average, than his peers.

Graduation tests. Many states mandate tests that have to be passed in order for the students to graduate with a high school diploma. A local school superintendent believes that after-school tutoring will improve the scores of students in his district on the state's graduation test. A tutor agrees to work with 15 students for a month before the superintendent will approach the school board about implementing an after-school tutoring program. The after-school tutoring program will be implemented if student scores increase by more than 20 points. The superintendent will test a hypothesis using $\alpha = 0.02$.

a) Write appropriate hypotheses (in words *and* in symbols).

$H_0: \mu_d = 20$ (The difference between the mean # of pts before and after the tutoring program is not more than 20)

$H_a: \mu > 20$

(The difference btw the mean # of pts before and after the tutoring program is more than 20)

b) In this context, which do you consider to be more serious – a Type I or a Type II error? Explain.

↳ Ho is true, but reject.

→ Concluding more than 20 when in reality not more than 20. School would spend money on a program that did not help.

Type 2: Ho is false but fail to reject.
↳ Concluding the program is ineffective when in reality it is effective.

not spending money on a program that works.

c) After this trial produced inconclusive results, the superintendent decided to test the after-school tutoring program again with another group of students. Describe two changes he could make in the trial to increase the power of the test, and explain the disadvantages of each.

↑

↑ α

↑ n.

↑ α: it could lead to adopting a tutoring program that actually doesn't help.

↑ n: trial cost would increase.

4. **Haircuts.** You need to find a new hair stylist and know that there are two terrific salons in your area, Hair by Charles and Curl Up & Dye. You want a really good haircut, but you do not want to pay too much for the cut. A random sample of costs for 10 different stylists was taken at each salon (each salon employs over 100 stylists).

a) Indicate what inference procedure you would use to see if there is a significant difference in the costs for haircuts at each salon. Check the appropriate assumptions and conditions and indicate whether you could or could nor proceed. (Do not do the actual test.)

• Random: randomly selected.

2-sample t-Test

• Independent Group: stylists from 2 different salons are definitely independent groups

• 10%: 10 < 10% of al possible stylists from each salon.

• Normality: we do not have data, so we do not know about this condition. we would proceed w/ caution

b) A friend tells you that he has heard that Curl Up & Dye is the more expensive salon.

i. Write hypotheses for your friend's claim.

H = Hair by Charles
C = Curl UP & Dye

$H_o: \mu_H - \mu_C = 0$

$H_a: \mu_H - \mu_C < 0$

ii. The following are computer outputs. Which output is the correct one to use for this test? Explain.

Output A:

```
Two-sample T for Hair by Charles vs Curl Up & Dye

                  N    Mean   StDev   SE Mean
Hair by Charles   10   22.10   6.33      2.0
Curl Up & Dye     10   26.00   4.81      1.5

Difference = mu (Hair by Charles) - mu (Curl Up & Dye)
Estimate for difference:    -3.90000
95% CI for difference:   (-9.22983, 1.42983)
T-Test of difference = 0 (vs not =): T-Value = -1.55   P-Value = 0.140
DF = 16
```

A b/c we are doing a 2-sample t-Test.

Output B:

```
Paired T for Hair by Charles - Curl Up & Dye

                  N       Mean     StDev    SE Mean
Hair by Charles   10    22.1000   6.3325    2.0025
Curl Up & Dye     10    26.0000   4.8074    1.5202
Difference        10   -3.90000   7.37036   2.33071

95% CI for mean difference: (-9.17244, 1.37244)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.67   P-Value =
0.129
```

← paired test.

iii. Use the appropriate computer output to make a conclusion about the hypothesis test based on the data. Make sure to state your conclusion in context.

The pvalue is 0.07, which is high, so we fail to reject Ho. There is insufficient evidence that curl UP & Dye is any more expensive on average than Hair by charles.