Stat 242 Quiz – Topics Drawn from Sections 5.5 and Chapter 3

What's Your Name? _____

In each of the following data problems there is a potential violation of the assumption of independence. Explain what the potential problem is in a sentence or two.

1. Researchers interested in learning the effects of speed limits on traffic accidents recorded the number of accidents per year for each of 10 consecutive years on roads in a state with speed limits of 90 km/h. They also recorded the number of accidents for the next 7 years on the same roads after the speed limit had been increased to 110 km/h. They conducted statistical analyses to compare the mean number of accidents per year when the speed limit was 90 km/h and the mean number of accidents per year when the speed limit was 110 km/h.

The same roads were measured multiple times at both speed limits, and those measurements are probably not independent of each other. For example, some roads may be more dangerous and may consistently have more accidents. Multiple measurements for the same road cannot be treated as independent. Additionally, the number of accidents may be more similar in consecutive years than in years that are far apart. Measurements in consecutive years may not be independent.

2. Researchers interested in investigating the effect of indoor pollution on respiratory health randomly selected houses in a particular city. Each house was monitored for nitrogen dioxide concentration and categorized as being either high or low on the nitrogen dioxide scale. Each member of the household was measured for respiratory health in terms of breathing capacity. They conducted statistical analyses to compare the mean breathing capacity score for people in houses with high nitrogen dioxide levels and the mean breathing capacity score for people in houses with low nitrogen dioxide levels.

The individuals within a given household may be related, and may therefore have similar respiratory health for genetic reasons; their measurements cannot be treated as independent.

(this page left intentionally blank)

Stat 242 Quiz – Topics Drawn from Sections 5.5 and Chapter 3

What's Your Name? _____

Researchers examined the time in minutes before an insulating fluid lost its insulating property when the fluid was exposed to each of two different voltages. They had eight samples of the fluid, 3 of which were randomly assigned to receive 26 kV of electricity and 5 of which were randomly assigned to receive 28 kV of electricity. The times until loss of insulating properties were skewed right with several outliers, so they performed a logarithmic transformation; after transformation the standard deviations within each group were similar.

The R code and output below shows the results of their analysis:

```
insulation %>%
  group_by(voltage) %>%
  summarize(
    mean(log_time)
  )
```

```
## # A tibble: 2 x 2
##   voltage `mean(log_time)`
##   <fct>              <dbl>
## 1 v26                 5.62
## 2 v28                 5.33
```

```
lm_fit <- lm(log_time ~ voltage, data = insulation)
fit.contrast(lm_fit, "voltage", c(1, -1), conf.int = 0.95)
```

```
##                     Estimate Std. Error   t value  Pr(>|t|)  lower CI upper CI
## voltage c=( 1 -1 ) 0.2945269   1.570726 0.1875101 0.8574411 -3.548901 4.137955
## attr(,"class")
## [1] "fit_contrast"
```

1. Interpret the estimated mean log times until loss of insulating properties calculated above in terms of what they say about a measure of the center of the distribution of times on the original data scale (in minutes).

You may use the following R output:

```
exp(5.62)
```

```
## [1] 275.8894
```

```
exp(5.33)
```

```
## [1] 206.438
```

It is estimated that in the population of all similar samples of this liquid, the median time until loss of insulating properties is about 276 minutes at a voltage of 26 kV, and about 206 minutes at a voltage of 28 kV

(see question 2 on the other side!)

2. The researchers calculated an estimate and a 95% confidence interval for the difference in mean log times. Interpret what the results say about the relationship between a measure of the center of the distribution of times for each group on the original data scale (in minutes). In your answer, include a description of the meaning of the phrase "95% confident".

You may use the following R output:

```
exp(0.29)
```

```
## [1] 1.336427
```

```
exp(-3.55)
```

```
## [1] 0.02872464
```

```
exp(4.14)
```

```
## [1] 62.80282
```

It is estimated that in the population of all similar samples of this liquid, the median time until loss of insulating properties is about 1.33 times longer at a voltage of 26kV than the median time at a voltage of 28kV. We are 95% confident that this multiplicative difference in median times until loss of insulating properties is between about 0.03 and 62.8. For 95% of samples, an interval calculated in this way would contain this multiplicative difference in medians for these groups.