Navid Morshed, Han Xu, Ryan Quigley
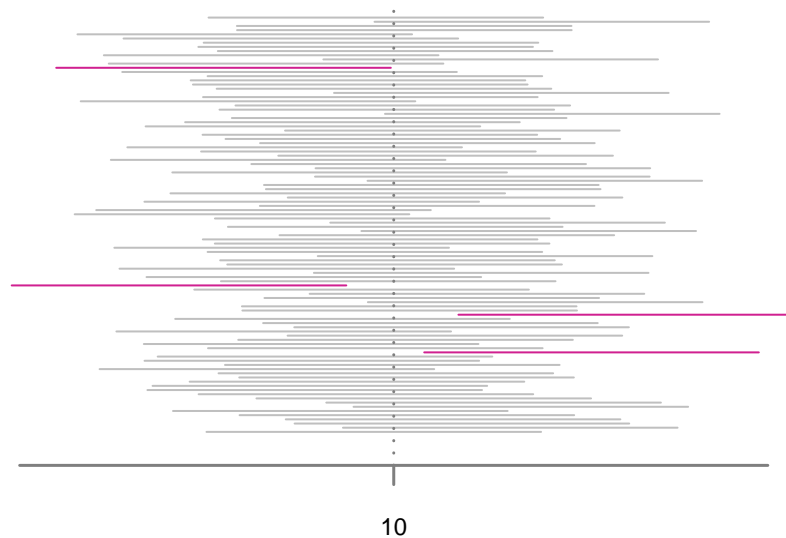MATH 267A
Section 2
Group A

Problem Set 10

1.

a) 4/100 intervals fail to capture the true mean. This is not surprising because sampling variation can cause the coverage of the confidence interval to differ slightly from 95% exactly. If we draw an infinite number of samples of the same size from the same population and calculate a confidence interval for each sample, the coverage of the intervals would be 95%; in other words, in the limit we expect the coverage of the intervals to approach 95%.

95% Confidence Intervals (100 samples)

10

4/100 fail to capture true mean

b) 21/100 intervals fail to capture the true mean. Again, sampling variation can cause the coverage of the confidence interval to differ slightly from 80% exactly, but we expect the coverage to approach 80% in the limit. The width of the 80% confidence intervals is narrower (2.34 compared to 3.58) because we must sacrifice confidence in order to gain higher accuracy.
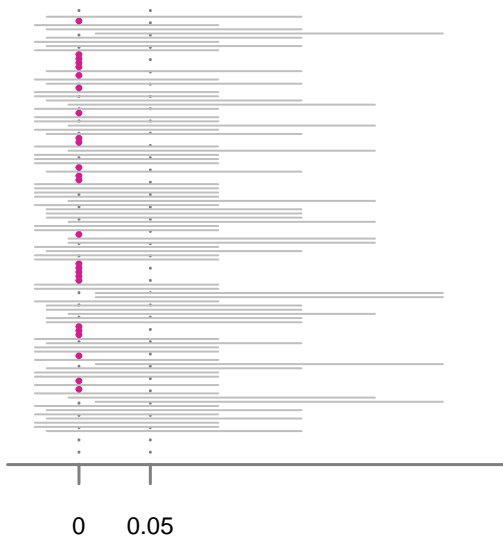
**80% Confidence Intervals (100 samples)**

10
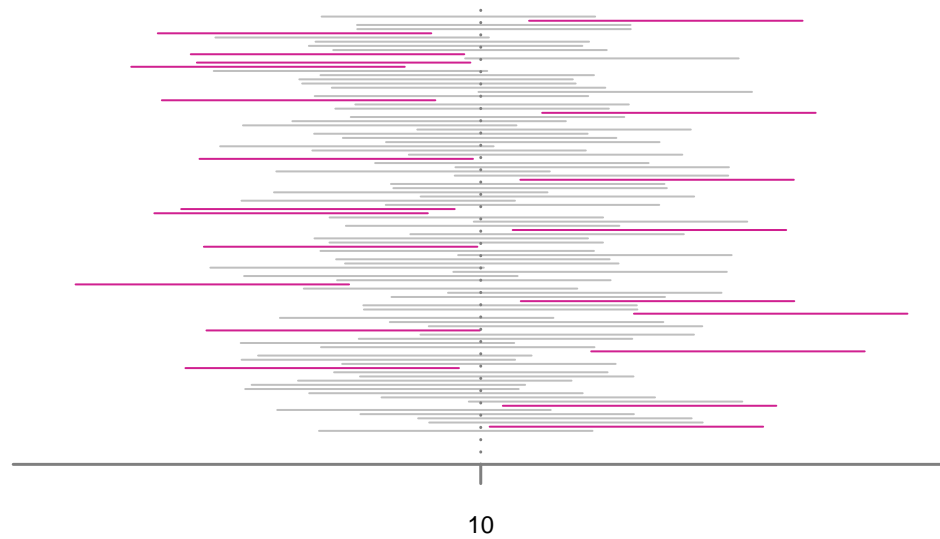
21/100 fail to capture true mean

c) It is worth noting that one of the conditions for normal approximation is not met: $p \cdot n = 0.05 \cdot 30 = 1.5 \ngeq 10$. 25% of Wald intervals fail to capture the true proportion, 5% of score intervals fail to capture the true proportion. There does not appear to be any benefit to using asymptotic interval estimators for binomial proportions. They produce intervals with bounds outside the parameter space, and they fail to calculate an interval when the number of successes in the sample is 0.

95% Wald Intervals (100 samples)   **95% Score Intervals (100 samples)**

0   0.05   0   0.05

25% fail to capture true proportion   5% fail to capture true proportion

2.
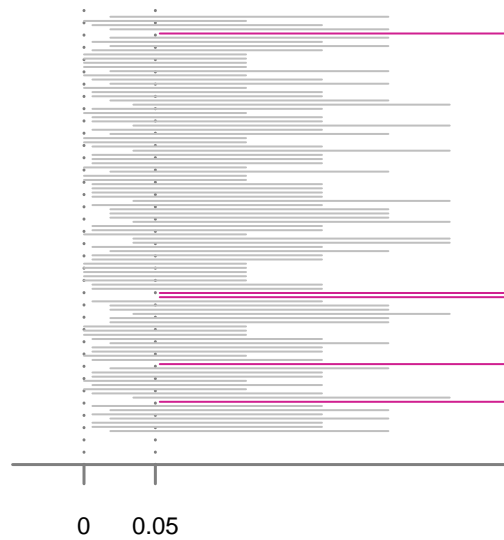
a) No discussion required

b) The 98% confidence interval for the average time between events, $1/\lambda$, is (6.33, 28.78)

c) In order to guarantee that the confidence interval contains the true value for the average time between events, the interval must be the entire support of the distribution: the support of the χ^2 distribution is $[0, \infty)$

```
ciExp(arv.time, alpha = 0)
```

```
## lcl ucl
##   0 Inf
```

3.

a) The null hypothesis: the previous survey results are still accurate and relevant. The alternative hypothesis: the old survey results are outdated and logging cars are actually being loaded more heavily in recent years.

$$H_0 : \mu \leq 140$$
$$H_a : \mu > 140$$

We want to determine whether or not to build the trestle to higher tolerances to accommodate heavier cars, so we are really only interested in the case where the average weight is greater than the hypothesized value. Thus, a one-sided test is appropriate.

In this context, type II error is a very serious problem. If we fail to reject the null hypothesis when in reality the average weight is greater than 140, then the engineering corps will not build the trestle to higher tolerances and the trestle will be at greater risk of structural failure. We can control type II error by making the rejection region larger, which can be achieved by choosing a smaller critical value. Determining that particular value should be left to the engineers, but for the sake of proceeding with the test we will choose a single number from the set of values associated with the alternative hypothesis. Say at a true average weight of 150 we want to make sure the type II error is at most 1%. We can calculate the critical value by solving the following equation for c:

$$t_{\beta,n-1} = \frac{c - \mu_a}{s/\sqrt{n}}$$
$$t_{0.01,19} = \frac{c - 150}{14.44/\sqrt{20}}$$
$$\implies c = 150 - 2.54 \cdot \frac{14.44}{\sqrt{20}}$$
$$= 141.83$$

3

Thus, our rejection region is $\bar{X} \geq 141.83$. In terms of quantiles, this correpondes to $t > \frac{141.83 - 140}{14.44/\sqrt{20}} = 0.57$

b) In order to justify performing a t-test, we must assume the underlying population is normally distributed. Assuming that is the case, the t-test is appropriate because we do not know the population standard deviation and our sample size is not large enough to justify a large-sample asymptotic test. The `t.test()` gives a t value of 1.97, which is significantly larger than our critical value, 0.57. Thus, we reject the null hypothesis and conclude that the average weight of loaded logging cars is greater than 140.

```
##
##  One Sample t-test
##
## data:  trains
## t = 1.9698, df = 19, p-value = 0.03181
## alternative hypothesis: true mean is greater than 140
## 95 percent confidence interval:
##  140.7771       Inf
## sample estimates:
## mean of x
##    146.36
```

c) The t-test from part (b) only allows us to advise the engineers to build the trestle to higher tolerance; however, it does not allow us to suggest a minimum threshold for the tolerance. We would need to construct a confidence interval in order to provide the engineers with suggested tolerance minimums.

d) There are a lot of unknown factors here. For example we don't have any knowledge of the cost factor. If the two bridges cost the same then build the strong one regardless of what our hypothesis test suggests. Our data is 10 years old and it's from a single year, so we don't know if this data valuable today. Will assumptions underlying this dataset be valuable in twenty years? Have the car weights changed over the past years? How will they change in the future? How has the climate changed over the last ten years? These are the questions that will affect our decision.

So our assumptions for part D is based on the fact that there is a significant difference between the prices of two bridges and that the weight of the cars, weight of the delivery product (if any), and the capacity the cars will stay the same for a long time. Also that the climate has remained the same for the past 10 years and will continue to do so for a long time.

According to the new information our data in no longer random since the cars were handpicked. So our result from the previous section is no longer valid. Our first natural approach should be to collect more data. If collecting more data is absolutely impossible and we are forced to work with this non-random data, then we can assume a larger critical value and adjust our calculations.