STAT 509 – Sections 4.4,4.8 – More Inference

• We can do inference (CIs, hypothesis tests) about parameters other than a population mean.

Confidence Interval for a Proportion

• Suppose our data tell us only whether each observation has a certain characteristic.
• We want to know how much of the population has that characteristic.
• The proportion (always between 0 and 1) of individuals with a characteristic is the same as the probability of a random individual having the characteristic.

Estimating proportion is equivalent to estimating the binomial probability *p*.

Point estimate of *p* is the <u>sample proportion</u>:

• Give every sampled individual a 1 (if it has the characteristic) or 0 (if it lacks it).

Note $\hat{p} = \frac{y}{n}$ **is a type of sample average (of 0's and 1's), so CLT tells us that when sample size is large, sampling distribution of** $\hat{p}$ **is approximately normal.**

For large *n*:

$100(1 - \alpha)\%$ CI for *p* is:

How large does *n* need to be?

Example 1: We wish to estimate the probability that a randomly selected part in a shipment will be defective. Take a random sample of 179 parts, and find 14 defective parts. Find a 95% CI for *p*.

Check:

Hypothesis Tests about a Population Proportion

We often wish to test whether a population proportion *p* equals a specified value.

Example 1 again: We wish to test whether the proportion of defective parts in a shipment is less than 0.10.

We test:

Recall: The sample proportion $\hat{p}$ is approximately $N\left(p, \sqrt{\frac{pq}{n}}\right)$ for large *n*, so our test statistic for testing H_0: $p = p_0$

has a standard normal distribution when H_0 is true (when *p* really is p_0).

Rules for one-tailed tests about population proportion

$H_0: p = p_0$		$H_0: p = p_0$
$H_a: p < p_0$	**or**	$H_a: p > p_0$

Test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

Rejection Region:	$z < -z_{\alpha}$	$z > z_{\alpha}$

Rules for two-tailed tests about population proportion

$H_0: p = p_0$

$H_a: p \neq p_0$

Test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

Rejection Region: $z < -z_{\alpha/2}$ **or** $z > z_{\alpha/2}$ **(both tails)**

Assumptions of test (need large sample):

Need:

Example 1:
Test H_0: $p = 0.10$ vs. H_a: $p < 0.10$ using $\alpha = .01$.

Take a random sample of 179 parts, and find 14 defective parts.

In R:
```
> prop.test(14,179, p=0.10, alternative="less",
correct=F)
```

Example 1(a): What if we had wanted to test whether the proportion of defective parts was <u>different from</u> 0.10?

In R:
```
> prop.test(14,179, p=0.10,
alternative="two.sided", correct=F)
```

Section 4.8 – Inference about Variances

<u>Confidence Interval for the Variance</u> σ^2 (or for s.d. σ)

Recall that if the data are normally distributed,

$\frac{(n-1)s^2}{\sigma^2}$ **has a χ^2 sampling distribution with (n – 1) d.f.**

This can be used to develop a (1 – α)100% CI for σ^2:

Note: This procedure is not robust! It is not appropriate if the data are not normal. Be sure to check the normality assumption!

• We can also derive a set of hypothesis tests (based on the χ^2 distribution) for testing whether the population variance equals some specified value.

Rules for one-tailed tests about population variance

$H_0: \sigma^2 = \sigma^2_0$ $\quad$ $H_0: \sigma^2 = \sigma^2_0$

$H_a: \sigma^2 < \sigma^2_0$ $\quad$ **or** $\quad$ $H_a: \sigma^2 > \sigma^2_0$

Test statistic: $$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Rejection Region: $\chi^2 < \chi^2_{n-1,1-\alpha}$ $\quad$ $\chi^2 > \chi^2_{n-1,\alpha}$

Rules for two-tailed tests about population variance

$H_0: \sigma^2 = \sigma^2_0$

$H_a: \sigma^2 \neq \sigma^2_0$

Test statistic: $$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Rejection Region: $\chi^2 < \chi^2_{n-1,1-\alpha/2}$ **or** $\chi^2 > \chi^2_{n-1,\alpha/2}$ **(both tails)**

Assumptions of test:

How to check this?

Example: A random sample of 10 lubricant containers yields s = 0.24585 liters, so s^2 =
(Assume normally distributed data)
Find 95% CI for σ^2.

95% CI for σ:

Testing whether the true variance is greater than 0.03:

• Recall that if we have independent samples from two normally distributed populations (having variances σ_1^2 and σ_2^2), then

$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$ **has an F sampling distribution with (n_1 – 1) numerator and (n_2 – 1) denominator d.f.**

• Therefore, if $\sigma_1^2 = \sigma_2^2$, then s_1^2 / s_2^2 has an F-distribution.

• Then a ratio of sample variances can serve as the test statistic for testing the hypotheses:

• Again, this procedure is not robust and is not appropriate unless both data sets are from normal populations.

Example: If we have two samples from normal populations, we can test for equal variances in R:

```
> Lu1 <- c(10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9,
10.4, 10.3, 9.8)
> Lu2 <- c(9.78, 9.79, 10.33, 9.91, 9.38, 10.09,
10.17, 9.44)
> qqnorm(Lu1) # checking normality assumption
> qqnorm(Lu2) # checking normality assumption
> var.test(Lu1, Lu2)
```