35. Confidence Intervals for Other Parameters

CI's for Normal Variance
IQ Test Example

CI's for Ratio of Two Normal Variances
Another IQ Test Example

CI's for a Bernoulli Proportion
Test Question Example

Confidence Intervals for Normal Variance

How much variability can we expect from observations from some system?

Now we'll get a CI for the variance of a normal distribution (instead of the mean).

Usual set-up:

$$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2).$$

Then recall that the distribution of the sample variance is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \frac{\sigma^2\chi^2(n-1)}{n-1},$$

i.e.,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

Using χ^2 quantiles and some algebra, we get

$$
\begin{aligned}
1-\alpha &= \Pr\left(\chi^2_{1-\frac{\alpha}{2},n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2},n-1}\right) \\
&= \Pr\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2},n-1}} \ge \frac{\sigma^2}{(n-1)S^2} \ge \frac{1}{\chi^2_{\frac{\alpha}{2},n-1}}\right) \\
&= \Pr\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right).
\end{aligned}
$$

So a $100(1-\alpha)\%$ CI for σ^2 is

$$\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}},\ \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right].$$

Remark: The confidence interval for σ^2 is directly proportional to the sample variance S^2.

Remark: This CI contains no reference to the unknown μ!

Meanwhile, a $100(1-\alpha)\%$ lower-CI for σ^2 is:

$$\frac{(n-1)S^2}{\chi^2_{\alpha,n-1}} \leq \sigma^2.$$

$100(1-\alpha)\%$ upper-CI for σ^2:

$$\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}.$$

Example: Suppose 25 people take an IQ test and that their scores are normally distributed.

If $S^2 = 100$, find a 95% upper-CI for the variance σ^2.

Looking up the χ^2 quantile, we get

$$\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}} = \frac{(24)(100)}{\chi^2_{0.05,24}} = \frac{2400}{13.85} = 173.3.$$

CI's for the Ratio of Variances of Two Normals

Which of two normal distributions is more variable?

Set-up:

$$X_1, X_2, \ldots, X_{n_x} \overset{\text{iid}}{\sim} \text{Nor}(\mu_x, \sigma_x^2) \quad \text{and}$$

$$Y_1, Y_2, \ldots, Y_{n_y} \overset{\text{iid}}{\sim} \text{Nor}(\mu_y, \sigma_y^2).$$

All X's and Y's are independent.

We'll get a CI for the *ratio* σ_x^2/σ_y^2.

Recall the distributions of the two sample variances:

$$S_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{n_x}(X_i - \bar{X})^2 \sim \frac{\sigma_x^2 \chi^2(n_x - 1)}{n_x - 1}, \quad \text{and}$$

$$S_y^2 = \frac{1}{n_y - 1}\sum_{i=1}^{n_y}(Y_i - \bar{Y})^2 \sim \frac{\sigma_y^2 \chi^2(n_y - 1)}{n_y - 1}.$$

Thus,

$$\frac{S_y^2/\sigma_y^2}{S_x^2/\sigma_x^2} \sim \frac{\chi^2(n_y - 1)/(n_y - 1)}{\chi^2(n_x - 1)/(n_x - 1)} \sim F(n_y - 1, n_x - 1).$$

Using F quantiles and some algebra, we get

$$\begin{aligned}
&1-\alpha \\
&= \Pr\left(F_{1-\frac{\alpha}{2},n_y-1,n_x-1} \le \frac{S_y^2/\sigma_y^2}{S_x^2/\sigma_x^2} \le F_{\frac{\alpha}{2},n_y-1,n_x-1}\right) \\
&= \Pr\left(\frac{S_x^2}{S_y^2}\,\frac{1}{F_{\frac{\alpha}{2},n_x-1,n_y-1}} \le \frac{\sigma_x^2}{\sigma_y^2} \le \frac{S_x^2}{S_y^2}\,F_{\frac{\alpha}{2},n_y-1,n_x-1}\right).
\end{aligned}$$

So a $100(1-\alpha)\%$ CI for σ_x^2/σ_y^2 is

$$\frac{\sigma_x^2}{\sigma_y^2} \in \left[\frac{S_x^2}{S_y^2}\,\frac{1}{F_{\frac{\alpha}{2},n_x-1,n_y-1}},\;\; \frac{S_x^2}{S_y^2}\,F_{\frac{\alpha}{2},n_y-1,n_x-1}\right].$$

Remark: The confidence interval for σ_x^2/σ_y^2 is proportional to the ratio of the sample variances, S_x^2/S_y^2.

Remark: This CI contains no reference to μ_x or μ_y.

Meanwhile, a $100(1-\alpha)\%$ lower-CI for σ_x^2/σ_y^2 is:

$$\frac{S_x^2}{S_y^2}\,\frac{1}{F_{\alpha,n_x-1,n_y-1}} \;\le\; \frac{\sigma_x^2}{\sigma_y^2}.$$

$100(1-\alpha)\%$ upper-CI for σ_x^2/σ_y^2:

$$\frac{\sigma_x^2}{\sigma_y^2} \;\le\; \frac{S_x^2}{S_y^2}\,F_{\alpha,n_y-1,n_x-1}.$$

Remark: If you want CI's for σ_y^2/σ_x^2, just flip all of the X's and Y's in the various CI's discussed above.

Example: Suppose 25 people take an IQ test A and 16 people take IQ test B. Assume all scores are normal and independent.

If $S_A^2 = 100$ and $S_B^2 = 70$, find a 95% upper-CI for the variance σ_A^2/σ_B^2.

Looking up the $F_{\alpha,n_B-1,n_A-1} = F_{0.05,15,24}$ quantile, we get

$$\frac{\sigma_A^2}{\sigma_B^2} \leq \frac{S_A^2}{S_B^2} F_{\alpha,n_B-1,n_A-1} = \frac{100}{70}(2.11) = 3.01.$$

Confidence Intervals for a Bernoulli Proportion

Suppose that $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$.

What probability of "success' can we expect from this distribution?

We'll get a CI for the proportion p of successes.

Since $\sum_{i=1}^{n} X_i \sim \text{Bin}(n, p)$, we know that

$$\bar{X} \sim \frac{\text{Bin}(n,p)}{n}.$$

Let's assume that n is "large", so we'll be able to use the Central Limit Theorem.

(If n isn't large, then we'll have to use nasty Binomial tables, which I don't want to deal with here!)

Note that

$$\mathsf{E}[\bar{X}] = \mathsf{E}[X_i] = p \text{ and}$$

$$\mathsf{Var}(\bar{X}) = \mathsf{Var}(X_i)/n = pq/n.$$

Then for large n, the CLT implies

$$\frac{\bar{X} - \mathsf{E}[\bar{X}]}{\sqrt{\mathsf{Var}(\bar{X})}} = \frac{\bar{X} - p}{\sqrt{pq/n}} \approx \mathsf{Nor}(0,1).$$

Now let's do something crazy and estimate pq by its MLE, $\bar{X}(1-\bar{X})$. This gives

$$\frac{\bar{X}-p}{\sqrt{\bar{X}(1-\bar{X})/n}} \approx \text{Nor}(0,1).$$

Then the "usual" algebra implies

$$\begin{aligned}
1-\alpha & \\
\approx\ & \Pr\left(-z_{\alpha/2} \le \frac{\bar{X}-p}{\sqrt{\bar{X}(1-\bar{X})/n}} \le z_{\alpha/2}\right) \\
=\ & \Pr\left(\bar{X}-z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \le p \le \bar{X}+z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right).
\end{aligned}$$

So an *approximate* 2-sided CI for p is

$$p \in \bar{X} \pm z_{\alpha/2} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}.$$

Similarly, an approx lower-CI is

$$\bar{X} - z_{\alpha} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \le p,$$

and an approx upper CI is

$$p \le \bar{X} + z_{\alpha} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}.$$

Example: The probability that a student correctly answers a certain test question is p.

Suppose a random sample of 100 students yields 40 correct answers to the question.

Find a 90% 2-sided CI for p.

$$p \in \bar{X} \pm z_{\alpha/2} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

$$= 0.4 \pm (1.645)\sqrt{(.4)(.6)/100} = 0.4 \pm 0.081.$$

The half-width of the 2-sided CI is

$$z_{\alpha/2}\sqrt{\frac{pq}{n}} \approx z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}.$$

How many observations should we take so that the half-length is $\le \epsilon$?

$$z_{\alpha/2}\sqrt{\frac{pq}{n}} \le \epsilon \iff n \ge (z_{\alpha/2}/\epsilon)^2 pq.$$

Of course, p and q are unknown. A *conservative* choice for n arises by maximizing $pq = 1/4$. Then we have

$$n \geq (z_{\alpha/2}/\epsilon)^2/4.$$

Of course, if we somehow make a *preliminary estimate* of p, we could use

$$n \geq (z_{\alpha/2}/\epsilon)^2\, \widehat{p}(1-\widehat{p}).$$