Lecture 17
Confidence Intervals

Review:

- The **sampling distribution of $\overline{\boldsymbol{x}}$** has mean of μ and standard deviation of $\frac{\sigma}{\sqrt{n}}$

- The **sampling distribution of $\widehat{\boldsymbol{p}}$** has a mean of p and a standard deviation of $\sqrt{\frac{p(1-p)}{n}}$
- The **standard error** of a statistic is just its standard deviation

μ - denotes the population mean

p - denotes the population proportion

σ – the population standard deviation

n – the sample size

Review 2

That interval is defined as $\pm 2 \times \mathrm{SE}$ from the mean:

The probability that $\bar{x}$ will be between $\mu - 2\,{}^{\sigma}/_{\sqrt{n}}$ and $\mu + 2\,{}^{\sigma}/_{\sqrt{n}}$ is approximately 0.95

$$P\left(\mu\ - 2\ \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 2\frac{\sigma}{\sqrt{n}}\right) \approx 0.95$$

The probability that $\hat{p}$ will be between $p - 2\sqrt{{}^{p(1-p)}/_{n}}$ and $p + 2\sqrt{{}^{p(1-p)}/_{n}}$ is approximately 0.95

$$P\left(p\ - 2\sqrt{\frac{p(1-p)}{n}} < \hat{p} < \mu + 2\sqrt{\frac{p(1-p)}{n}}\right) \approx 0.95$$


0.95
μ − 2σ/√n
μ
μ + 2σ/√n
x̄

Practice: Crooked Casino

- A crooked casino uses loaded dice at all of their Craps tables to improve their earnings. The table to left gives the probability distribution for the sum of roll of two die for a pair of fair dice (denoted X_{fair}) and for a pair of loaded dice (denoted X_{loaded})

X	$P(X_{\text{loaded}})$	$P(X_{\text{fair}})$
2	0.03	0.028
3	0.06	0.056
4	0.08	0.083
5	0.10	0.111
6	0.14	0.139
7	0.28	0.167
8	0.12	0.139
9	0.10	0.111
10	0.05	0.083
11	0.02	0.056
12	0.03	0.028

Practice: Crooked Casino

- Suppose a gambler at the casino is suspects that the casino is using loaded dice so he observes the proportion of “sums of 7” rolled in the next 30 turns at the Craps table. He computes the proportion of rolls that summed to 7 to be 0.33

- Assuming the dice are fair, Compute the interval that has a probability of approximately 0.95 of containing estimated proportion of rolls that sum to 7

$$\hat{p} \approx N\left(0.167, \sqrt{\frac{0.167(1-0.167)}{30}}\right) \approx N(0.167, 0.068)$$

$$P(0.167 - 2 \times 0.068 < \hat{p} < 0.167 + 2 \times 0.068) = 0.95$$

$$P(0.031 < \hat{p} < 0.303) = 0.95$$

X	$P(X_{\text{loaded}})$	$P(X_{\text{fair}})$
2	0.03	0.028
3	0.06	0.056
4	0.08	0.083
5	0.10	0.111
6	0.14	0.139
7	0.28	0.167
8	0.12	0.139
9	0.10	0.111
10	0.05	0.083
11	0.02	0.056
12	0.03	0.028

Practice: Crooked Casino

- Suppose a gambler at the casino is suspects that the casino is using loaded dice so he observes the proportion of "sums of 7" rolled in the next 30 turns at the Craps table. He computes the proportion of rolls that summed to 7 to be 0.33

- Assuming the dice are fair, what is the probability of observing a proportion greater than the gamblers estimate?

$$SE(\hat{p}) = \sqrt{\frac{0.167(1 - 0.167)}{30}} = 0.068$$

$$z = \frac{0.33 - 0.167}{0.068} = 1.89$$

$$P(z > 2.39) = 1 - P(z \leq 2.39) = 0.0084$$

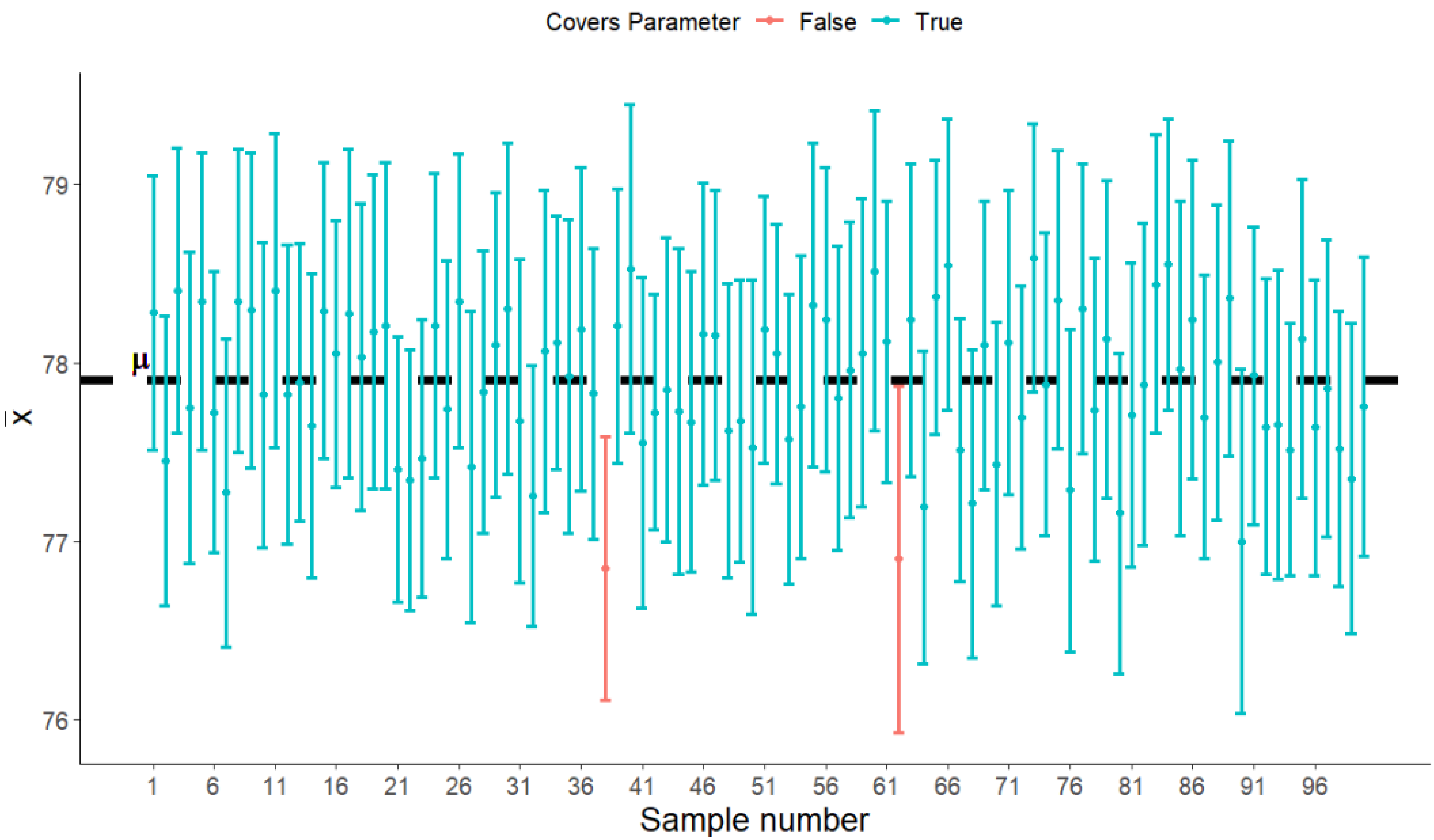
X	$P(X_{\text{loaded}})$	$P(X_{\text{fair}})$
2	0.03	0.028
3	0.06	0.056
4	0.08	0.083
5	0.10	0.111
6	0.14	0.139
7	0.28	0.167
8	0.12	0.139
9	0.10	0.111
10	0.05	0.083
11	0.02	0.056
12	0.03	0.028

Types of Estimation

There are two types of estimation

1. **Point estimation** - is estimation of the value of a parameter with the value of a statistic (i.e estimating μ with $\bar{x}$ or p with $\hat{p}$)

2. **Interval estimation** - is the estimation of the value of a parameter with an interval of values. The device we will be using for interval estimation is a *confidence* interval.



Covers Parameter
False
True
79
78
77
76
μ
X̄
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
Sample number

Confidence Interval for $\bar{x}$

With a little algebra we can show that

$$P\left(\mu - 2\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 2\frac{\sigma}{\sqrt{n}}\right) \approx 0.95$$

Then

$$P\left(\bar{x} - 2\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2\frac{\sigma}{\sqrt{n}}\right) \approx 0.95$$

The **Confidence Interval for** $\bar{x}$ is

$$\bar{x} \pm 2\frac{\sigma}{\sqrt{n}} \rightarrow \left(\bar{x} - 2\frac{\sigma}{\sqrt{n}}, \bar{x} + 2\frac{\sigma}{\sqrt{n}}\right)$$

Has a probability of approximately 0.95 of containing μ

we need to replace all parameters in the equations above with their respective estimates because the parameters are unknown

Confidence Interval for $\hat{p}$

- We can use the same approach to derive a confidence interval for a sample proportion

$$\hat{p} \pm 2\sqrt{\frac{p(1-p)}{n}} \rightarrow \left(\hat{p} - 2\sqrt{\frac{p(1-p)}{n}}, \hat{p} + 2\sqrt{\frac{p(1-p)}{n}}\right)$$

Has a probability of approximately 0.95 of containing p

Example

Example: Let x be whether or not polymerase chain reaction (PCR) test is successful under certain circumstances. Suppose we conduct 100 tests to produce a sample of 100 observations of x and observe that the PCR test was successful on 90 of those 100 observations. What is the confidence interval for estimating the probability of a successful PCR test?

Example: $\bar{x}$

- A teacher is estimating the heights of college students at the university of Idaho. From a sample of 50 students, the teacher estimates the average height to be 71.4 inches (about 5 foot 11 inches) with a variance of about 23 inches. Compute the 99% confidence interval for the mean height of college students

$$n = 50$$
$$\bar{x} = 71.4 \text{ inches}$$
$$s = \sqrt{23} \approx 4.8 \text{ inches}$$

$$SE(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{4.8}{\sqrt{50}} = 0.67$$