Inference for other estimators

Inference for other estimators

- The sample mean is not the only point estimate for which the sampling distribution is nearly normal. For example, the sampling distribution of sample proportions is also nearly normal when n is sufficiently large.
- An important assumption about point estimates is that they are *unbiased*, i.e. the sampling distribution of the estimate is centered at the true population parameter it estimates.
  - That is, an unbiased estimate does not naturally over or underestimate the parameter. Rather, it tends to provide a "good" estimate.
  - The sample mean is an example of an unbiased point estimate, as are each of the examples we introduce in this section.
- Some point estimates follow distributions other than the normal distribution, and some scenarios require statistical techniques that we haven't covered yet – we will discuss these at the end of this section.

Confidence intervals for nearly normal point estimates

A confidence interval based on an unbiased and nearly normal point estimate is

$$point\ estimate \pm z^{\star} SE$$

where $z^{\star}$ is selected to correspond to the confidence level, and SE represents the standard error.

Remember that the value $z^{\star} SE$ is called the *margin of error*.

One of the earliest examples of behavioral asymmetry is a preference in humans for turning the head to the right, rather than to the left, during the final weeks of gestation and for the first 6 months after birth. This is thought to influence subsequent development of perceptual and motor preferences. A study of 124 couples found that 64.5% turned their heads to the right when kissing. The standard error associated with this estimate is roughly 4%. Which of the below is false?

(a) The 95% confidence interval for the percentage of kissers who turn their heads to the right is roughly $64.5\% \pm 4\%$.

(b) A higher sample size would yield a lower standard error.

(c) The margin of error for a 95% confidence interval for the percentage of kissers who turn their heads to the right is roughly 8%.

(d) The 99.7% confidence interval for the percentage of kissers who turn their heads to the right is roughly $64.5\% \pm 12\%$.

Güntürkün, O. (2003) Adult persistence of head-turning asymmetry. *Nature*. Vol 421.

One of the earliest examples of behavioral asymmetry is a preference in humans for turning the head to the right, rather than to the left, during the final weeks of gestation and for the first 6 months after birth. This is thought to influence subsequent development of perceptual and motor preferences. A study of 124 couples found that 64.5% turned their heads to the right when kissing. The standard error associated with this estimate is roughly 4%. Which of the below is false?

(a) *The 95% confidence interval for the percentage of kissers who turn their heads to the right is roughly $64.5\% \pm 4\%$.*

(b) A higher sample size would yield a lower standard error.

(c) The margin of error for a 95% confidence interval for the percentage of kissers who turn their heads to the right is roughly 8%.

(d) The 99.7% confidence interval for the percentage of kissers who turn their heads to the right is roughly $64.5\% \pm 12\%$.

Güntürkün, O. (2003) Adult persistence of head-turning asymmetry. *Nature*. Vol 421.

Hypothesis testing for nearly normal point estimates

The third National Health and Nutrition Examination Survey collected body fat percentage (BF%) and gender data from 13,601 subjects ages 20 to 80. The average BF% for the 6,580 men in the sample was 23.9, and this value was 35.0 for the 7,021 women. The standard error for the difference between the average men and women BF%s was 0.114. Do these data provide convincing evidence that men and women have different average BF%s. You may assume that the distribution of the point estimate is nearly normal.

Hypothesis testing for nearly normal point estimates

The third National Health and Nutrition Examination Survey collected body fat percentage (BF%) and gender data from 13,601 subjects ages 20 to 80. The average BF% for the 6,580 men in the sample was 23.9, and this value was 35.0 for the 7,021 women. The standard error for the difference between the average men and women BF%s was 0.114. Do these data provide convincing evidence that men and women have different average BF%s. You may assume that the distribution of the point estimate is nearly normal.

1. Set hypotheses
2. Calculate point estimate
3. Check conditions
4. Draw sampling distribution, shade p-value
5. Calculate test statistics and p-value, make a decision

1. The null hypothesis is that men and women have equal average BF%, and the alternative is that these values are different.

$$H_0 : \mu_{men} = \mu_{women} \qquad H_A : \mu_{men} \neq \mu_{women}$$

1. The null hypothesis is that men and women have equal average BF%, and the alternative is that these values are different.

$$H_0 : \mu_{men} = \mu_{women} \qquad H_A : \mu_{men} \neq \mu_{women}$$

2. The parameter of interest is the average difference in the population means of BF%s for men and women, and the point estimate for this parameter is the difference between the two sample means:

$$\bar{x}_{men} - \bar{x}_{women} = 23.9 - 35.0 = -11.1$$

1. The null hypothesis is that men and women have equal average BF%, and the alternative is that these values are different.

$$H_0 : \mu_{men} = \mu_{women} \qquad H_A : \mu_{men} \neq \mu_{women}$$

2. The parameter of interest is the average difference in the population means of BF%s for men and women, and the point estimate for this parameter is the difference between the two sample means:

$$\bar{x}_{men} - \bar{x}_{women} = 23.9 - 35.0 = -11.1$$

3. We are assuming that the distribution of the point estimate is nearly normal (we will discuss details for checking this condition in the next chapter, however given the large sample sizes, the normality assumption doesn't seem unwarranted).

4. The sampling distribution will be centered at the null value ($\mu_{men} - \mu_{women} = 0$), and the p-value is the area beyond the observed difference in sample means in both tails (lower than -11.1 and higher than 11.1).
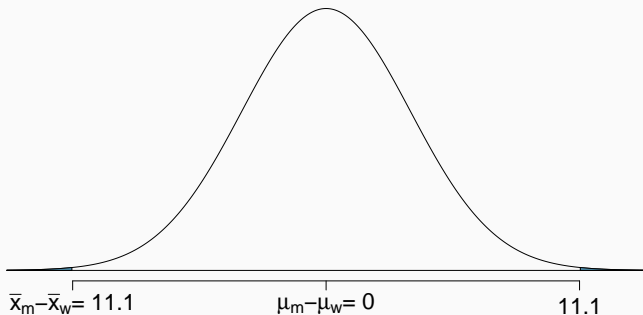
$\bar{x}_m - \bar{x}_w = 11.1$ $\mu_m - \mu_w = 0$ 11.1

5. The test statistic is computed as the difference between the point estimate and the null value (-11.1 - 0 = -11.1), scaled by the standard error.

$$Z = \frac{11.1 - 0}{0.114} = 97.36$$

The Z score is huge! And hence the p-value will be tiny, allowing us to reject H_0 in favor of H_A.

5. The test statistic is computed as the difference between the point estimate and the null value (-11.1 - 0 = -11.1), scaled by the standard error.

$$Z = \frac{11.1 - 0}{0.114} = 97.36$$

The Z score is huge! And hence the p-value will be tiny, allowing us to reject H_0 in favor of H_A.

These data provide convincing evidence that the average BF% of men and women are different.

Non-normal point estimates

- We may apply the ideas of confidence intervals and hypothesis testing to cases where the point estimate or test statistic is not necessarily normal. There are many reasons why such a situation may arise:
  - the sample size is too small for the normal approximation to be valid;
  - the standard error estimate may be poor; or
  - the point estimate tends towards some distribution that is not the normal distribution.
- For each case where the normal approximation is not valid, our first task is always to understand and characterize the sampling distribution of the point estimate or test statistic. Next, we can apply the general frameworks for confidence intervals and hypothesis testing to these alternative distributions.

- Statistical tools rely on the following two main conditions:
  - *Independence* A random sample from less than 10% of the population ensures independence of observations. In experiments, this is ensured by random assignment. If independence fails, then advanced techniques must be used, and in some such cases, inference may not be possible.
  - *Sample size and skew* For example, if the sample size is too small, the skew too strong, or extreme outliers are present, then the normal model for the sample mean will fail.
- Whenever conditions are not satisfied for a statistical technique:
  1. Learn new methods that are appropriate for the data.
  2. Consult a statistician.
  3. ~~Ignore the failure of conditions.~~ This last option effectively invalidates any analysis and may discredit novel and interesting findings.

All else held equal, will the p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$

(b) $n = 10,000$

All else held equal, will the p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$

(b) $n = 10,000$

All else held equal, will the p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$

(b) $n = 10,000$

Suppose $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, and $H_A : \mu > 49.5$.

All else held equal, will the p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$

(b) $n = 10,000$

Suppose $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, and $H_A : \mu > 49.5$.

$$Z_{n=100} \quad = \quad \frac{50 - 49.5}{\frac{2}{\sqrt{100}}}$$

All else held equal, will the p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$

(b) $n = 10,000$

Suppose $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, and $H_A : \mu > 49.5$.

$$Z_{n=100} = \frac{50 - 49.5}{\frac{2}{\sqrt{100}}} = \frac{50 - 49.5}{\frac{2}{10}} = \frac{0.5}{0.2} = 2.5, \quad \text{p-value} = 0.0062$$

All else held equal, will the p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$

(b) $n = 10,000$

Suppose $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, and $H_A : \mu > 49.5$.

$$Z_{n=100} = \frac{50 - 49.5}{\frac{2}{\sqrt{100}}} = \frac{50 - 49.5}{\frac{2}{10}} = \frac{0.5}{0.2} = 2.5, \quad \textit{p-value} = 0.0062$$

$$Z_{n=10000} = \frac{50 - 49.5}{\frac{2}{\sqrt{10000}}}$$

All else held equal, will the p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$

(b) $n = 10,000$

Suppose $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, and $H_A : \mu > 49.5$.

$$Z_{n=100} = \frac{50 - 49.5}{\frac{2}{\sqrt{100}}} = \frac{50 - 49.5}{\frac{2}{10}} = \frac{0.5}{0.2} = 2.5, \quad \textit{p-value} = 0.0062$$

$$Z_{n=10000} = \frac{50 - 49.5}{\frac{2}{\sqrt{10000}}} = \frac{50 - 49.5}{\frac{2}{100}} = \frac{0.5}{0.02} = 25, \quad \textit{p-value} \approx 0$$

All else held equal, will the p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$

(b) $n = 10,000$

Suppose $\bar{x} = 50$, $s = 2$, $H_0 : \mu = 49.5$, and $H_A : \mu > 49.5$.

$$Z_{n=100} = \frac{50 - 49.5}{\frac{2}{\sqrt{100}}} = \frac{50 - 49.5}{\frac{2}{10}} = \frac{0.5}{0.2} = 2.5, \quad \text{p-value} = 0.0062$$

$$Z_{n=10000} = \frac{50 - 49.5}{\frac{2}{\sqrt{10000}}} = \frac{50 - 49.5}{\frac{2}{100}} = \frac{0.5}{0.02} = 25, \quad \text{p-value} \approx 0$$

As n increases - SE ↓, Z ↑, p-value ↓

Test the hypothesis $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ for the following 8 samples. Assume $\sigma = 2$.

$\bar{x}$	10.05	10.1	10.2
$n = 30$	$p - value = 0.45$	$p - value = 0.39$	$p - value = 0.29$
$n = 5000$	$p - value = 0.39$	$p - value = 0.0002$	$p - value = 0$

Test the hypothesis $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ for the following 8 samples. Assume $\sigma = 2$.

$\bar{x}$	10.05	10.1	10.2
$n = 30$	$p - value = 0.45$	$p - value = 0.39$	$p - value = 0.29$
$n = 5000$	$p - value = 0.39$	$p - value = 0.0002$	$p - value = 0$

Test the hypothesis $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ for the following 8 samples. Assume $\sigma = 2$.

$\bar{x}$	10.05	10.1	10.2
$n = 30$	$p - value = 0.45$	$p - value = 0.39$	$p - value = 0.29$
$n = 5000$	$p - value = 0.04$	$p - value = 0.0002$	$p - value = 0$

Test the hypothesis $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ for the following 8 samples. Assume $\sigma = 2$.

$\bar{x}$	10.05	10.1	10.2
$n = 30$	$p - value = 0.45$	$p - value = 0.39$	$p - value = 0.29$
$n = 5000$	$p - value = 0.04$	$p - value = 0.0002$	$p - value = 0$

Test the hypothesis $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ for the following 8 samples. Assume $\sigma = 2$.

$\bar{x}$	10.05	10.1	10.2
$n = 30$	$p - value = 0.45$	$p - value = 0.39$	$p - value = 0.29$
$n = 5000$	$p - value = 0.04$	$p - value = 0.0002$	$p - value = 0$

Test the hypothesis $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ for the following 8 samples. Assume $\sigma = 2$.

$\bar{x}$	10.05	10.1	10.2
$n = 30$	$p-value = 0.45$	$p-value = 0.39$	$p-value = 0.29$
$n = 5000$	$p-value = 0.04$	$p-value = 0.0002$	$p-value = 0$

Test the hypothesis $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ for the following 8 samples. Assume $\sigma = 2$.

$\bar{x}$	10.05	10.1	10.2
$n = 30$	$p - value = 0.45$	$p - value = 0.39$	$p - value = 0.29$
$n = 5000$	$p - value = 0.04$	$p - value = 0.0002$	$p - value \approx 0$

Test the hypothesis $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ for the following 8 samples. Assume $\sigma = 2$.

$\bar{x}$	10.05	10.1	10.2
$n = 30$	$p - value = 0.45$	$p - value = 0.39$	$p - value = 0.29$
$n = 5000$	$p - value = 0.04$	$p - value = 0.0002$	$p - value \approx 0$

When n is large, even small deviations from the null (small effect sizes), which may be considered practically insignificant, can yield statistically significant results.

Statistical vs. practical significance

- Real differences between the point estimate and null value are easier to detect with larger samples.
- However, very large samples will result in statistical significance even for tiny differences between the sample mean and the null value (*effect size*), even when the difference is not practically significant.
- This is especially important to research: if we conduct a study, we want to focus on finding meaningful results (we want observed differences to be real, but also large enough to matter).
- The role of a statistician is not just in the analysis of data, but also in planning and design of a study.

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of." – R.A. Fisher