5 Point and Interval Estimation for a Single Sample

5.1 Point Estimation

5.2 Large-Sample Confidence Intervals for a Population Mean

5.3 Confidence Intervals for Proportions

5.4 Small-Sample Confidence Intervals for a Population Mean

An October 3, 2013, Gallup poll said U.S. unemployment was 7.9% in September. The fine print said, “results ... are based on telephone interviews conducted Sept. 1-30 with a random sample of 30,628 adults” and “with 95% confidence ... the margin of sampling error is ±1%.” Today we study how to make ________________, so that we can understand news stories like this one.

5.1 Point Estimation

Vocabulary:

- A **parameter** is a numerical summary of a ____________
  e.g. $p = \frac{\# \text{unemployed U.S. workers}}{\# \text{U.S. workers}}$ = proportion of U.S. workers who are unemployed
  e.g. $\mu, \sigma$, population minimum

- A **statistic** is a numerical summary of a ____________
  e.g. $\hat{p} = 7.9% = \frac{\# \text{unemployed workers in sample}}{\# \text{workers in sample}}$ = proportion of sample who are unemployed
  e.g. $\bar{X}, s_X$, sample minimum

- A **point estimate** (§5.1) is a ____________ used to estimate a ____________
  e.g. $\hat{p} = 7.9\%$ is a point estimate for $p$
  e.g. $\bar{X}$ and $s_X$ are point estimates for ______ and ______

- A **confidence interval** (§5.2) consists of a point estimate, an associated ________________, and a ________________
  e.g. $7.9\% \pm 1\% = (6.9\%, 8.9\%)$ is a 95% confidence interval for $p$ (§5.3)

A statistic $\hat{\theta}$ used to estimate a parameter $\theta$ should be ____________ and ____________.

- $\hat{\theta}$’s accuracy is measured by its **bias**, $\mu_{\hat{\theta}} - \theta$, the difference between its mean and the parameter.
  e.g. Find the bias of each estimator:
  - $X_1$ (from a population with $\mu$ and $\sigma$) as an estimator of $\mu$ has bias ____________
  - $\bar{X}$ as an estimator of $\mu$ has bias ____________
  - a sample minimum as an estimator of the population minimum has bias ____________
• $\hat{\theta}$’s precision is measured by its variance, $\sigma_{\hat{\theta}}^2$.

e.g. Find the variance of each estimator:

- $\sigma_{X_1}^2 = \underline{\text{__________}}$
- $\sigma_{X}^2 = \underline{\text{__________}}$

• The mean squared error of $\hat{\theta}$, $\text{MSE}_{\hat{\theta}} = \mu_{(\hat{\theta} - \theta)^2}$, is a measure of both accuracy and precision:

\[
\text{MSE}_{\hat{\theta}} = \mu_{(\hat{\theta} - \theta)^2} \\
= \mu_{\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2} + (0) \\
= \mu_{\hat{\theta}^2 - 2\mu_{\hat{\theta}}\theta + \theta^2} + (\mu_{\hat{\theta}^2} + \mu_{\theta}^2 - 2\mu_{\hat{\theta}}^2) \\
= (\mu_{\hat{\theta}^2} - 2\mu_{\hat{\theta}}\theta + \theta^2) + (\mu_{\hat{\theta}^2} - 2\mu_{\hat{\theta}}^2 + \mu_{\theta}^2) \\
= (\mu_{\hat{\theta}^2} - \theta)^2 + \mu_{\hat{\theta}^2 - 2\mu_{\hat{\theta}}\theta + \theta^2} \\
= (\mu_{\hat{\theta}^2} - \theta)^2 + (\hat{\theta} - \mu_{\hat{\theta}})^2 \\
= (\mu_{\hat{\theta}^2} - \theta)^2 + \sigma_{\hat{\theta}}^2 \\
= \underline{\text{_____________________}} + \underline{\text{_____________________}}
\]

The MSE of an unbiased estimator is just its ____________.

e.g. Considering $\bar{X}$ as an estimator for $\mu$, $\text{MSE}_{\bar{X}} = \underline{\text{_____________________}}$

e.g (p. 176 #3) $X_1$ and $X_2$ are independent, each with unknown mean $\mu$ and known $\sigma^2 = 1$.

a. (homework) Let $\hat{\mu}_1 = \frac{X_1 + X_2}{2}$. Find the bias, variance, and mean squared error of $\hat{\mu}_1$.

b. Let $\hat{\mu}_2 = \frac{X_1 + 2X_2}{3}$. Find the bias, variance, and mean squared error of $\hat{\mu}_2$.

c. (homework) Let $\hat{\mu}_3 = \frac{X_1 + X_2}{4}$. Find the bias, variance, and mean squared error of $\hat{\mu}_3$.

A point estimate alone isn’t very useful. Reporting it with it’s MSE is useful, but it’s more common to report a ______________ around a point estimate.
5.2 Large-Sample Confidence Intervals for a Population Mean

Derivation of Confidence Interval for Unknown \( \mu \) from \( \bar{X} \)

If \( X_1, \ldots, X_n \) is a large sample from a population with mean \( \mu \) and standard deviation \( \sigma \), then the CLT says that \( \bar{X} \sim N(\mu, \sigma^2/n) \) (approximately).

Here we construct an interval around \( \bar{X} \) which contains \( \mu \) for a proportion \( 1 - \alpha \) of random samples, where \( \alpha \in (0, 1) \). \( 100(1 - \alpha) \) is the confidence level of the interval.

Let \( z_{\alpha/2} \) be the \( z \)-score cutting off a right tail area of \( \alpha \) from \( N(0,1) \) (draw).

e.g. For the conventional confidence level 95%, \( \alpha = \) \( \) and \( z_{\alpha/2} = \) \( \)

Then \( P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha \) (draw). Unstandardize using \( Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \) to get

\[
P(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha,
\]

which we solve in two ways:

- for \( \bar{X} \) in the middle: \( P(\mu - z_{\alpha/2}\sigma/\sqrt{n} < \bar{X} < \mu + z_{\alpha/2}\sigma/\sqrt{n}) = 1 - \alpha \) (see picture)
- for \( \mu \): \( P(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n} < \mu < \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}) = 1 - \alpha \) (see picture)

For this \( \bar{x} \), \( \mu \) is \( \) the confidence interval. This happens with probability \( \) .

\[
\bar{X} \sim N(\mu, \sigma^2/n)
\]

For this \( \bar{x} \), \( \mu \) is \( \) the confidence interval. This happens with probability \( \) .
That is, \( \bar{X} \pm z_{\alpha/2} \sigma \sqrt{n} = \bar{X} \pm \frac{z_{\alpha/2} \sigma}{\sqrt{n}} \) contains \( \mu \) for a proportion \( 1 - \alpha \) of random samples. It’s the 100\%(1 - \alpha) confidence interval for \( \mu \).

We usually don’t know \( \sigma \), the population standard deviation. For a large sample, approximate it by ________________

e.g. Make a 95% confidence interval for the mean height \( \mu \) of UW men from a random sample of 30 men’s heights.

\[
1 - \alpha = \square \implies \alpha = \square \implies z_{\alpha/2} = \square
\]

\( \bar{X} = \square \)

\( s = \square \)

\( \bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} = \square \)

With what probability does our interval contain \( \mu \)? _____________

**How Confidence Intervals Behave**

- \( \bar{X} \pm \square \frac{s}{\sqrt{n}} \) is a 68% confidence interval for \( \mu \)

- \( \bar{X} \pm 1.96 \frac{s}{\sqrt{n}} \) is a 95% confidence interval for \( \mu \)

- \( \bar{X} \pm \square \frac{s}{\sqrt{n}} \) is a 99% confidence interval for \( \mu \)

- \( \bar{X} \pm \square \frac{s}{\sqrt{n}} \) is a 99.7% confidence interval for \( \mu \)

We want high confidence and a small margin of error, but the margin is \( z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \), which gets smaller when \( z_{\alpha/2} \) gets smaller, which corresponds to \( (1 - \alpha) \) getting smaller too. Extreme cases are that we can have confidence approaching 100\% as the margin approaches ________________, or we can have confidence approaching ______________ as the margin approaches 0.

**Choosing the Sample Size**

Good news is that the margin also gets smaller as _________________. For a desired margin of error \( m \), we can find the required sample size:

\[
m = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \implies
\]

(Use \( \sigma \approx s \) in the usual case where we don’t know \( \sigma \).)
§5.2 Example

e.g. (p. 187 #7) In a sample of 100 boxes of a certain type, the average compressive strength was 6230 N, and the standard deviation was 221 N.

a. Find a 95% confidence interval for the mean compressive strength.

b. Find a 99% confidence interval for the mean compressive strength.

c. An engineer claims that the mean strength is between 6205 and 6255 N. With what level of confidence can this statement be made?

d. Approximately how many boxes must be sampled so that a 95% confidence interval will specify the mean to within ±25 N?

e. Approximately how many boxes must be sampled so that a 99% confidence interval will specify the mean to within ±25 N?