TEXAS McCombs

The University of Texas at Austin

McCombs School of Business

Test 2 Review

Betsy Greenberg

What's on Test 2?

Chapters 6-8 and 10, 11

Skip the following topics:

- Chapter 6 - Uniform, Geometric, and Poisson Models Binomial calculations except for mean and standard deviation
- Chapter 7 - Continuity correction, Uniform and Exponential Models
- Chapter 8 - Stratified, Cluster, Multistage, and Systematic samples
- Chapter 10 - don't skip anything
- Chapter 11 - Bootstrapping

Random Variables

- Mean

  $\mu_X = x_1P_1 + x_2P_2 + \ldots + x_kP_k$

- Variance

  ${\sigma_X}^2 = (x_1 - \mu_X)^2P_1 + (x_2 - \mu_X)^2P_2 + \ldots + (x_k - \mu_X)^2P_k$

- Standard Deviation

  $\sigma_X = \sqrt{{\sigma_X}^2}$

Adding and subtracting random variables:

$E(X \pm Y) = E(X) \pm E(Y)$

$Var(X \pm Y) = V(X) + Var(Y)$ (if X and Y are **independent**)

Bernoulli
B is the number of successes from one trial
where p is the probability of success
$E[B] = p$ and $Var(B) = p(1 - p)$

Binomial
X is the number of successes from n independent trials
where p is the probability of success
$E[X] = np$ and $Var(X) = np(1 - p)$

Sample Proportion
$\hat{p} = \frac{X}{n}$ is the proportion of successes from n independent trials
where p is the probability of success
$\mathrm{E}[\hat{p}] = \frac{E[X]}{n} = p$ and $Var(\hat{p}) = \frac{Var(X)}{n^2} = \frac{p(1-p)}{n}$

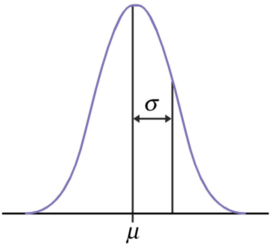
We did not cover Uniform, Geometric, or Poisson dist

Binomial Distributions

- The total number of observations *n* **is fixed** in advance.
- The outcomes of all *n* observations are statistically **independent.**
- Each observation falls into just one of 2 categories: **success** and **failure.**
- Same **probability of success** for each trial

We did not cover calculating Binomial probabilities

Normal Distribution

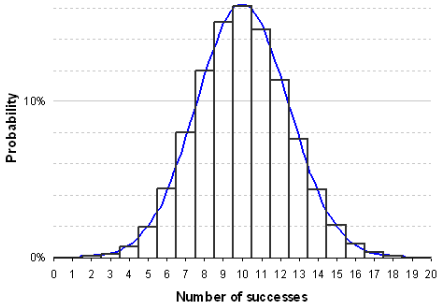


- 68 - 95 - 99.7 rule
- Standardized observations
- Normal distribution calculations:
  - Find areas - Forward Calculations
  - Find a value when given an area - Backward Calculations

Calculate probabilities for the Normal distribution:

- `pnorm(x, mean, standard_dev)` finds areas to the left of x
- `qnorm(probability, mean, standard_dev)` finds the value with the specified probability to the left
- Use `hist` and `qqnorm` to check if data is normal
- For calculations about the sample mean, $\bar{x}$, use $\frac{s}{\sqrt{n}}$ for standard_dev

Normal approximation for the Binomial



If $np \geq 10$ and $n(1-p) \geq 10$ the Binomial distribution is approximately Normal with

$$\mu = np$$
$$\sigma = \sqrt{np(1-p)}$$

We did not cover Uniform or Exponential distributions

Sampling

- Populations
  Parameters describe populations
- Samples
  Statistics describe samples
- We hope that sample data is representative of the population
- Sampling variability - sample to sample differences (also called sampling error)
- Non-sampling errors - due to voluntary response, non-response, poorly worded questions, etc.

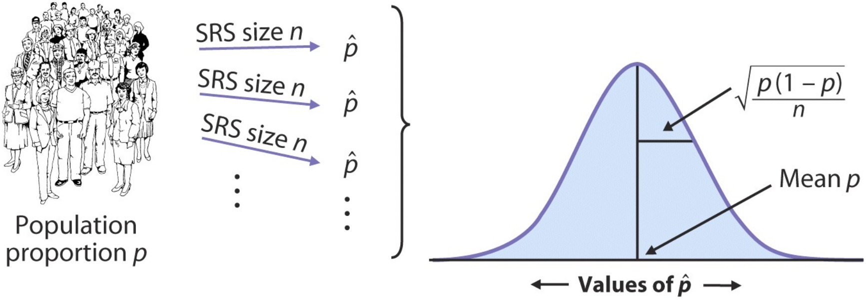
Big Ideas

- Sample - Examine a Part of the whole
- Randomize - to avoid bias
- The Sample Size is what matters

Possible causes of Bias

- Voluntary response samples
- Under coverage
- Non-response
- Behavior or appearance of interviewer
- Poorly worded questions
- Interviewer fabrications

Sampling Distribution of $\hat{p}$



The mean of the sampling distribution is p

The standard deviation is $\sqrt{\frac{p(1-p)}{n}}$

Since p is unknown,

we use the standard error instead which is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

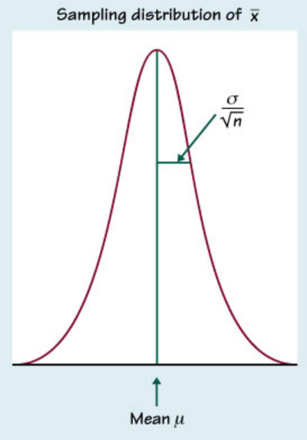
Sampling Distribution of $\bar{x}$

- If x_i has mean μ and standard deviation σ,
  $\bar{x} = \sum \frac{x_1 + \cdots + x_n}{n}$ has mean μ

- If the x_is are independent, the standard deviation $= \frac{\sigma}{\sqrt{n}}$

- Since σ is typically unknown, it will be estimated with the sample standard deviation, s

- The **standard error is** $\frac{s}{\sqrt{n}}$



Checklist for Confidence Intervals:

- **Independence:**

  The sampled values must be independent of each other.
- **10% condition:**

  Sample size is less than 10% of the population size.
- **Randomization:**

  The sample is a simple random sample from the relevant population.
- **Sample size condition:**

  Success\failure condition for proportions

  Normal population assumption

Specific sample size conditions:

- **Confidence Interval for proportion:**
  Both $n\hat{p}$ and $n(1-\hat{p})$ are at least 10.

- **CI for means:**
  n is greater than both 10(skewness)^2 and 10|kurtosis-3|

`library(moments)`

`skewness`, `kurtosis`

Confidence Intervals

A confidence interval has the form:

$$\text{Estimate} \;\pm\; \text{Margin of Error}$$

$$\text{Estimate} \;\pm\; (z^* \;\; or \;\; t^*) \times \text{Standard Error (SE)}$$

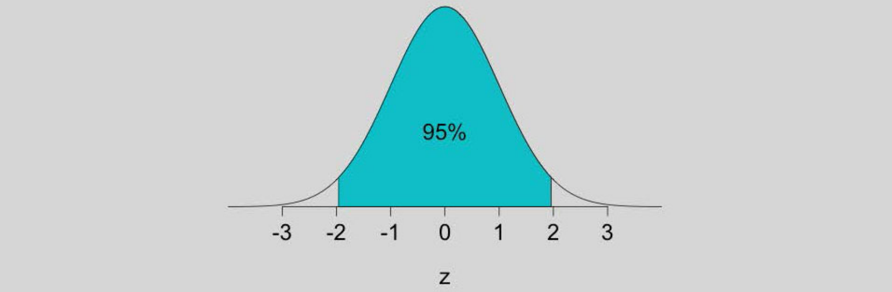
Confidence Intervals

- The confidence level *C*, shows how confident we are that the procedure will catch the true population parameter.
- The procedures give confidence intervals that C% of the time will include the true population parameter
- Type of Problems
  - Proportions
  - Means

Confidence Intervals for: Population Proportion, *p*

- Estimate: $\hat{p} = \frac{X}{n} = \frac{\text{Number of successes}}{\text{number of trials}}$
- Standard Error: $SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- Critical Value: z^*
- Margin of Error: $z^* \, SE = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Finding $\pm z^*$ for 95% confidence

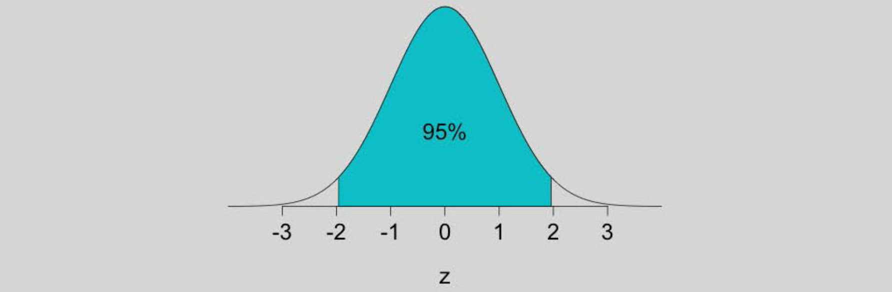


`qnorm`(0.025) and `qnorm`(0.975)

Confidence Intervals for: Population Mean

- Estimate: $\bar{x}$ (`mean`)
- Standard Error: $SE = \frac{s}{\sqrt{n}}$
  where $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$ (`sd`)
- Degrees of freedom: $k = n - 1$
- Critical Value: t^* for a distribution $t(k)$
- Margin of Error: $t^*\,SE = t^*\,\frac{s}{\sqrt{n}}$

Finding $\pm t^*$ for 95% confidence and df = n - 1



`qt`(0.025, df) and `qt`(0.975, df)

R functions

`ci.prop` for confidence interval for poroportion

`t.test` for inference for means

(ci.prop function and data will be given in an .RData file)

Determining Sample Size

Since the margin of error is $z^* \sqrt{p(1-p)/n}$,

we can find the sample size using

$$n = \left(\frac{z^*}{\text{Margin of Error}} \right)^2 p(1-p)$$

Obtain an estimate for p using an earlier sample (since we have to choose n before collecting data)

Use $p = 0.5$ to be conservative

ALWAYS ROUND UP for sample size calculations

Margin of Error $= t^* \frac{s}{\sqrt{n}}$

Determining Sample Size

$$n = \left(\frac{t^* s}{\text{Margin of Error}} \right)^2$$

For a study about μ with 95% coverage, find the sample size using

$$n \approx \left(\frac{2s}{\text{Margin of Error}} \right)^2$$