Designing the Study

PSY 300

Housekeeping

- Being healthy and happy is more important than any assignment—if you're having trouble finishing something, email me

Housekeeping

- Being healthy and happy is more important than any assignment—if you're having trouble finishing something, email me
- Today's lecture *is* relevant for the midterm.

Housekeeping

- Being healthy and happy is more important than any assignment—if you're having trouble finishing something, email me
- Today's lecture *is* relevant for the midterm.
- Read and understand the textbook—I won't spend much time on it in lecture, but I will expect you to know it. If you have questions, ask the email list! Or come to office hours.

Housekeeping

- Being healthy and happy is more important than any assignment—if you're having trouble finishing something, email me
- Today's lecture *is* relevant for the midterm.
- Read and understand the textbook—I won't spend much time on it in lecture, but I will expect you to know it. If you have questions, ask the email list! Or come to office hours.
- ***Do not be scared of the math. Nothing you will have to do goes beyond plugging stuff into other stuff and doing rote arithmetic, or letting the computer do it.*** If you get stuck, ask the email list, or talk to me or the TAs!

Housekeeping

- Being healthy and happy is more important than any assignment—if you're having trouble finishing something, email me
- Today's lecture *is* relevant for the midterm.
- Read and understand the textbook—I won't spend much time on it in lecture, but I will expect you to know it. If you have questions, ask the email list! Or come to office hours.
- ***Do not be scared of the math. Nothing you will have to do goes beyond plugging stuff into other stuff and doing rote arithmetic, or letting the computer do it.*** If you get stuck, ask the email list, or talk to me or the TAs!
- The important (and non-trivial!) bit is getting the *concepts* right—focus on that.

Housekeeping

- Being healthy and happy is more important than any assignment—if you're having trouble finishing something, email me
- Today's lecture *is* relevant for the midterm.
- Read and understand the textbook—I won't spend much time on it in lecture, but I will expect you to know it. If you have questions, ask the email list! Or come to office hours.
- ***Do not be scared of the math. Nothing you will have to do goes beyond plugging stuff into other stuff and doing rote arithmetic, or letting the computer do it.*** If you get stuck, ask the email list, or talk to me or the TAs!
- The important (and non-trivial!) bit is getting the *concepts* right—focus on that.
- Khan Academy and Wikipedia are great on these topics!

Today: Sampling and hypothesis testing review

1. Drawing samples from distributions
2. Calculating estimators of interest (e.g. means)
3. Calculating measures of dispersion (e.g. standard error)
4. Testing hypotheses about estimators using the standard error

Concept review

- The outcomes we study are ***random variables*** that have ***distributions***. They are defined in ***populations***, and we usually study them by taking ***samples*** from the population.

Concept review

- The outcomes we study are ***random variables*** that have ***distributions***. They are defined in ***populations***, and we usually study them by taking ***samples*** from the population.
- Example:
    - Random variable: height of humans
    - Distribution: probably approximately normal
    - Population: all humans
    - Example of a sample: representative sample of 1000 people from each country.

Population means and sample means

Quantities when sampling from a distribution of a random variable, x:

Population means and sample means

Quantities when sampling from a distribution of a random variable, x:

- Population mean μ: The "expectation" of x (formally μ is defined as $\mu = E[x]$). This is the *true* mean of the distribution. We never observe it—we can only estimate it from a sample. Example: true average height of humans.

Population means and sample means

Quantities when sampling from a distribution of a random variable, x:

- Population mean μ: The "expectation" of x (formally μ is defined as $\mu = E[x]$). This is the *true* mean of the distribution. We never observe it—we can only estimate it from a sample. Example: true average height of humans.
- Sample mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$. This is the sample approximation of the population mean. Example: average height of a random sample of humans.

Population standard deviation and sample standard deviation

- Population standard deviation $\sigma = \sqrt{Var(x)} = \sqrt{E[(x-\mu)^2]}$: measure of variance/dispersion in the population. E.g. variation in the height of humans. The units are the same as those of the variable of interest (e.g., inches).

Population standard deviation and sample standard deviation

- Population standard deviation $\sigma = \sqrt{Var(x)} = \sqrt{E[(x-\mu)^2]}$: measure of variance/dispersion in the population. E.g. variation in the height of humans. The units are the same as those of the variable of interest (e.g., inches).
- Sample standard deviation: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$. This is the sample approximation of the population standard deviation. E.g. variation of height in a random sample of humans. We also write "SD" sometimes.

Standard error

- The standard error is the **standard deviation of the sampling distribution** of a statistic.

Standard error

- The standard error is the **standard deviation of the sampling distribution** of a statistic.
- E.g., the standard error of the mean is the standard deviation of the sampling distribution of the mean

Standard error

- The standard error is the **standard deviation of the sampling distribution** of a statistic.
- E.g., the standard error of the mean is the standard deviation of the sampling distribution of the mean
- If we repeatedly sample several units from the population, and calculate the mean each time, this procedure itself will give us a distribution that has its own mean and standard deviation. That standard deviation is called the standard error of the mean.

Standard error

- This is different from the standard deviation *of the population* itself.
  - Importantly, it depends on sample size: the larger the sample, the smaller the standard error = the larger the sample, the more precisely we estimate the mean.

Standard error

- This is different from the standard deviation *of the population* itself.
  - Importantly, it depends on sample size: the larger the sample, the smaller the standard error = the larger the sample, the more precisely we estimate the mean.
- Small standard errors are good—they give us precision. We are always looking for ways to decrease our standard errors.

Standard error

- This is different from the standard deviation *of the population* itself.
  - Importantly, it depends on sample size: the larger the sample, the smaller the standard error = the larger the sample, the more precisely we estimate the mean.
- Small standard errors are good—they give us precision. We are always looking for ways to decrease our standard errors.
  - Easiest way: increase the sample size.

Standard error

- This is different from the standard deviation *of the population* itself.
  - Importantly, it depends on sample size: the larger the sample, the smaller the standard error = the larger the sample, the more precisely we estimate the mean.
- Small standard errors are good—they give us precision. We are always looking for ways to decrease our standard errors.
  - Easiest way: increase the sample size.
- Standard errors are defined for sampling distributions of *statistics*. The *mean* is an example of a statistic. Another statistic that we will often calculate the standard error for is a treatment effect, i.e. a difference in outcomes between treated and control conditions. These statistics are also called *estimators*.

Standard error

- Standard error of the mean: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Standard error

- Standard error of the mean: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.
- How do we estimate the standard error of the mean, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$? Recall that we don't observe the population standard deviation σ directly. Solution: We use the sample standard deviation.

Standard error

- Standard error of the mean: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.
- How do we estimate the standard error of the mean, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$? Recall that we don't observe the population standard deviation σ directly. Solution: We use the sample standard deviation.
- Thus, the standard error of the mean is estimated as: $s_{\bar{x}} = \frac{s}{\sqrt{n}}$.

Standard error

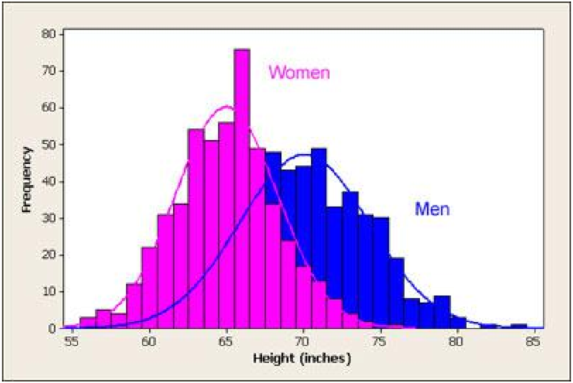
- Standard error of the mean: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.
- How do we estimate the standard error of the mean, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$? Recall that we don't observe the population standard deviation σ directly. Solution: We use the sample standard deviation.
- Thus, the standard error of the mean is estimated as: $s_{\bar{x}} = \frac{s}{\sqrt{n}}$.
- We also write "*SE*" for the standard error sometimes, and "SEM" for the standard error of the mean.

Standard error

- Standard error of the mean: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.
- How do we estimate the standard error of the mean, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$? Recall that we don't observe the population standard deviation σ directly. Solution: We use the sample standard deviation.
- Thus, the standard error of the mean is estimated as: $s_{\bar{x}} = \frac{s}{\sqrt{n}}$.
- We also write "*SE*" for the standard error sometimes, and "SEM" for the standard error of the mean.
- So:

$$SE = \frac{SD}{\sqrt{n}}$$

The Normal Distribution



Null Hypothesis Significance Testing (NHST)

- The average height of women in the US is 65 inches.

Null Hypothesis Significance Testing (NHST)

- The average height of women in the US is 65 inches.
- Question: Are WNBA basketball players taller than 65 inches on average?

Null Hypothesis Significance Testing (NHST)

- The average height of women in the US is 65 inches.
- Question: Are WNBA basketball players taller than 65 inches on average?
- Sample: 100 randomly chosen WNBA players

Null Hypothesis Significance Testing (NHST)

- The average height of women in the US is 65 inches.
- Question: Are WNBA basketball players taller than 65 inches on average?
- Sample: 100 randomly chosen WNBA players
- Null hypothesis H_0: $\mu = 65$

Null Hypothesis Significance Testing (NHST)

- The average height of women in the US is 65 inches.
- Question: Are WNBA basketball players taller than 65 inches on average?
- Sample: 100 randomly chosen WNBA players
- Null hypothesis H_0: $\mu = 65$
- Alternative hypothesis H_1: $\mu \neq 65$

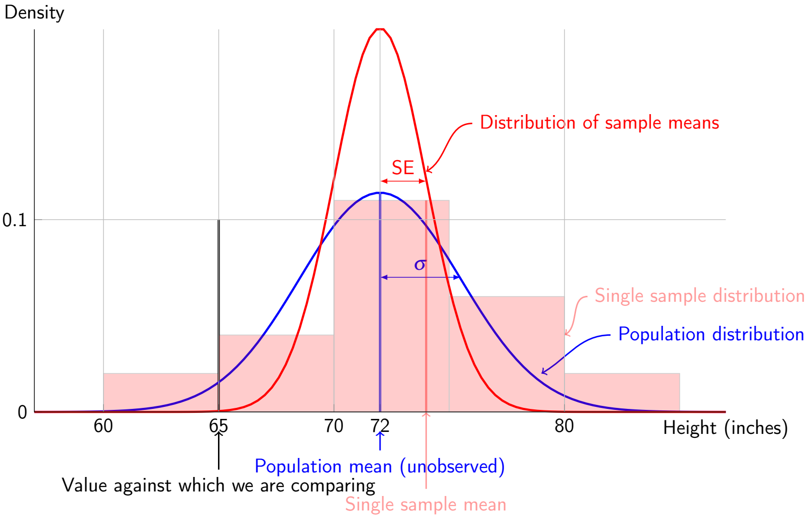
Null Hypothesis Significance Testing (NHST)

- The average height of women in the US is 65 inches.
- Question: Are WNBA basketball players taller than 65 inches on average?
- Sample: 100 randomly chosen WNBA players
- Null hypothesis H_0: $\mu = 65$
- Alternative hypothesis H_1: $\mu \neq 65$
  - (Note this is not $\mu > 65$, which would imply a one-tailed test. You probably learned that one-tailed tests are the thing to do when you have a directional hypothesis. In practice it's frowned upon and rarely done.)

Null Hypothesis Significance Testing (NHST)

- The average height of women in the US is 65 inches.
- Question: Are WNBA basketball players taller than 65 inches on average?
- Sample: 100 randomly chosen WNBA players
- Null hypothesis H_0: $\mu = 65$
- Alternative hypothesis H_1: $\mu \neq 65$
  - (Note this is not $\mu > 65$, which would imply a one-tailed test. You probably learned that one-tailed tests are the thing to do when you have a directional hypothesis. In practice it's frowned upon and rarely done.)
- Significance level: $\alpha = 0.05$
  - Probability of rejecting the null if the null is true ("false positive" or "Type I error")

Sampling from a Normal Distribution



Sampling from a Normal Distribution

- When we repeatedly draw samples from a population and calculate the mean each time, we gradually obtain the distribution of sample means. Most means will be pretty close to the population mean; some will be farther away. Over time they will converge ever closer to the population mean.

Sampling from a Normal Distribution

- When we repeatedly draw samples from a population and calculate the mean each time, we gradually obtain the distribution of sample means. Most means will be pretty close to the population mean; some will be farther away. Over time they will converge ever closer to the population mean.
- The larger our individual samples are, the tighter the distribution of means is centered around the population mean. For example, if a sample only has 5 observations, the sample mean can be very different from the population mean—e.g. if we happen to draw 5 very tall people. In contrast, when we draw more people, the sample mean will be closer to the population mean—e.g. if we draw 1000 people, their average height won't be too different from that of the population.

Sampling from a Normal Distribution

- Magically, the sampling distribution of means will be normally distributed, even if the underlying variable is not! (Central Limit Theorem)

Sampling from a Normal Distribution

- Magically, the sampling distribution of means will be normally distributed, even if the underlying variable is not! (Central Limit Theorem)
- We don't have to keep drawing samples: we can estimate the mean and SE directly from one sample.
  - Mean $\bar{x} = \frac{1}{n}\sum_1^n x_i$
  - Standard error: $SE = \frac{SD}{\sqrt{n}}$

Significance testing

- To determine if the observed sample mean is *significantly* different from a value of interest (e.g. 65 inches), we ask if it lies in the tails of the distribution under the null; i.e., if it's very unlikely if the null is true. If the observed sample mean lies in the most extreme 5% of the distribution under the null, we say the sample mean is "significantly different at the 5 percent level" from the value of interest.

Significance testing

- To determine if the observed sample mean is *significantly* different from a value of interest (e.g. 65 inches), we ask if it lies in the tails of the distribution under the null; i.e., if it's very unlikely if the null is true. If the observed sample mean lies in the most extreme 5% of the distribution under the null, we say the sample mean is "significantly different at the 5 percent level" from the value of interest.
  - In large samples, this is the same as asking if the observed sample mean is more than 1.96 SE away from the value of interest.

“Distribution under the Null”

- The “distribution under the null” is the sampling distribution we would be drawing our samples from ***if*** the null hypothesis were true.

“Distribution under the Null”

- The “distribution under the null” is the sampling distribution we would be drawing our samples from ***if*** the null hypothesis were true.
- It has the mean specified by the null hypothesis; in our example, 65 inches.

"Distribution under the Null"

- The "distribution under the null" is the sampling distribution we would be drawing our samples from ***if*** the null hypothesis were true.
- It has the mean specified by the null hypothesis; in our example, 65 inches.
- It has the standard error observed in the data.

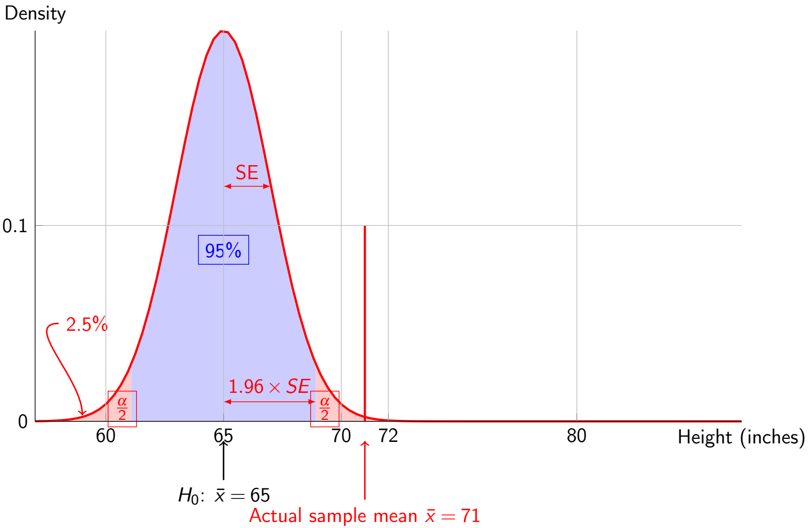
“Distribution under the Null”

- The “distribution under the null” is the sampling distribution we would be drawing our samples from ***if*** the null hypothesis were true.
- It has the mean specified by the null hypothesis; in our example, 65 inches.
- It has the standard error observed in the data.
- We can then ask: *If* the null hypothesis were true ($\mu = 65$), what is the probability that we would obtain a sample mean at least as extreme as the one we observe (e.g. $\bar{x} = 72$)?

"Distribution under the Null"

- The "distribution under the null" is the sampling distribution we would be drawing our samples from ***if*** the null hypothesis were true.
- It has the mean specified by the null hypothesis; in our example, 65 inches.
- It has the standard error observed in the data.
- We can then ask: *If* the null hypothesis were true ($\mu = 65$), what is the probability that we would obtain a sample mean at least as extreme as the one we observe (e.g. $\bar{x} = 72$)?
  - **This is the definition of the p-value**

"Distribution under the Null"



Significance testing

- How do we test if the estimator is significantly different from zero? Compute the probability of obtaining an estimator *at least as extreme as this one* if the null were true.

Significance testing

- How do we test if the estimator is significantly different from zero? Compute the probability of obtaining an estimator *at least as extreme as this one* if the null were true.
- In practice, we first compute the difference we are testing: $71 - 65$. This is our ***estimator***.

Significance testing

- How do we test if the estimator is significantly different from zero? Compute the probability of obtaining an estimator *at least as extreme as this one* if the null were true.
- In practice, we first compute the difference we are testing: $71 - 65$. This is our ***estimator***.
  - Note this is kind of a re-arranged H_0: "Is the sample mean (71) larger than 65?" is the same as "Is the difference between 71 and 65 greater than zero?"

Significance testing

- How do we test if the estimator is significantly different from zero? Compute the probability of obtaining an estimator *at least as extreme as this one* if the null were true.
- In practice, we first compute the difference we are testing: $71 - 65$. This is our ***estimator***.
    - Note this is kind of a re-arranged H_0: "Is the sample mean (71) larger than 65?" is the same as "Is the difference between 71 and 65 greater than zero?"
- We then test if the estimator is significantly different from zero. We do this by ***dividing the estimator by the standard error***; this gives us the "**test statistic**". It measures how many standard errors the estimator is away from zero.

Significance testing

- Because in most cases the test statistic follows the t-distribution, it's also called the *t*-statistic:

$$t = \frac{estimator}{SE}$$

- This is the same one-sample t-test you encountered last year! Recall (or not ;)) that you learned there: $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$. We've now just called $\bar{x}-\mu$ the estimator, and $\frac{s}{\sqrt{n}}$ the standard error.

Significance testing

- Because in most cases the test statistic follows the t-distribution, it's also called the *t*-statistic:

$$t = \frac{estimator}{SE}$$

- This is the same one-sample t-test you encountered last year! Recall (or not ;)) that you learned there: $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$. We've now just called $\bar{x} - \mu$ the estimator, and $\frac{s}{\sqrt{n}}$ the standard error.
- Many other test can be expressed in the same form: e.g. independent-sample t-tests, repeated-measures t-tests, interactions in ANOVA... We will do this later in the course.

Significance testing

- We then determine the p-value associated with the t-statistic we obtain, using the degrees of freedom of the t-test $(n-1)$.
  - $p(\bar{x} \geq 71|H_0) < 0.05$: Reject H_0
  - $p(\bar{x} \geq 71|H_0) \geq 0.05$: Do not reject H_0

Significance testing: Example

- Average height in sample: $\bar{x} = 71$

Significance testing: Example

- Average height in sample: $\bar{x} = 71$
- Standard error of sample: $SE = 2$

Significance testing: Example

- Average height in sample: $\bar{x} = 71$
- Standard error of sample: $SE = 2$
- Estimator: $71 - 65$

Significance testing: Example

- Average height in sample: $\bar{x} = 71$
- Standard error of sample: $SE = 2$
- Estimator: $71 - 65$
- Test statistic: $t = \frac{71-65}{2} = 3.5$. So the sample mean is 3.5 SE away from the comparison value. (Equivalently: the estimator is 3.5 SE away from zero.)

Significance testing: Example

- Average height in sample: $\bar{x} = 71$
- Standard error of sample: $SE = 2$
- Estimator: $71 - 65$
- Test statistic: $t = \frac{71-65}{2} = 3.5$. So the sample mean is 3.5 SE away from the comparison value. (Equivalently: the estimator is 3.5 SE away from zero.)
- Calculate the *p*-value using a table for the *t*-statistic. We had 100 observations here, so degrees of freedom = 99. In *R*: "1 - pt(3.5, df = 99)"

Significance testing: Example

- Average height in sample: $\bar{x} = 71$
- Standard error of sample: $SE = 2$
- Estimator: $71 - 65$
- Test statistic: $t = \frac{71-65}{2} = 3.5$. So the sample mean is 3.5 SE away from the comparison value. (Equivalently: the estimator is 3.5 SE away from zero.)
- Calculate the *p*-value using a table for the *t*-statistic. We had 100 observations here, so degrees of freedom = 99. In *R*: "1 - pt(3.5, df = 99)"
- The *p*-value is 0.026, so the difference is statistically significant at the 5% level. We reject H_0.

Significance testing: Example

- Average height in sample: $\bar{x} = 71$
- Standard error of sample: $SE = 2$
- Estimator: $71 - 65$
- Test statistic: $t = \frac{71-65}{2} = 3.5$. So the sample mean is 3.5 SE away from the comparison value. (Equivalently: the estimator is 3.5 SE away from zero.)
- Calculate the p-value using a table for the t-statistic. We had 100 observations here, so degrees of freedom = 99. In *R*: "1 - pt(3.5, df = 99)"
- The p-value is 0.026, so the difference is statistically significant at the 5% level. We reject H_0.
- Conclusion: "Our results suggest that the average height of WNBA basketball players is greater than that of the average woman in the US."