P values and Hypothesis testing for 1 mean or proportion

Is our sample evidence consistent with a particular hypothesis?

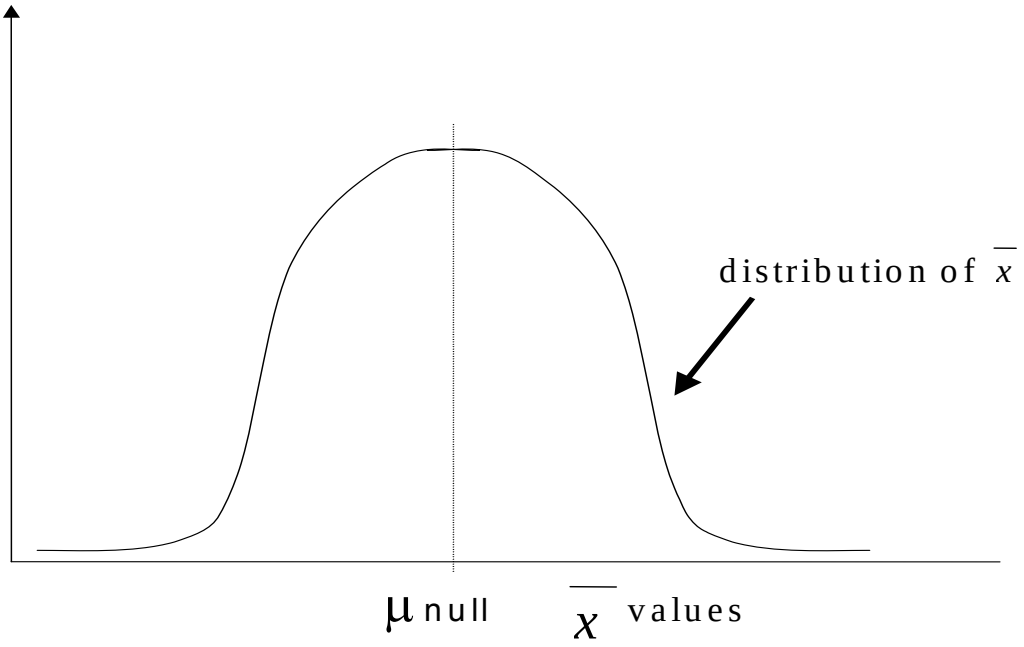
Null hypothesis- hypothesized value of population parameter

Alternative Hypothesis- everything but value of null hypothesis.

Do single female heads of households in Washington State make, on average, the 2006 federal poverty rate for a family of three?

Null hypothesis: μ= \$16,600 [This we hold to be true until falsified]

Alternative hypotheses: $\mu \neq$\$16,600 [Everything but the null hypothesis]



Do single female heads of households in Washington State make, on average, the poverty rate for a family of three ($16,600)?

In the 2006 WA state population survey we have a subsample of 1,167 female-headed households who reported an average income of $36,572 with a standard deviation of $29,311.

$$\bar{x} = \$36{,}572 \qquad SE = \frac{S}{\sqrt{n}} = \frac{29{,}311}{\sqrt{1167}} = 858$$

How far is the sample mean from the hypothesized population mean?

Find test statistic (like Z score):

$$z = \frac{\bar{x} - \mu_0}{SE} = \frac{36{,}572 - 16{,}600}{29{,}311/\sqrt{1167}} = \frac{19{,}972}{858} = 23.28$$

How likely is that you'd get a sample mean so far away from the population mean?

Probability you'd get a sample mean so far (or further) from the population mean

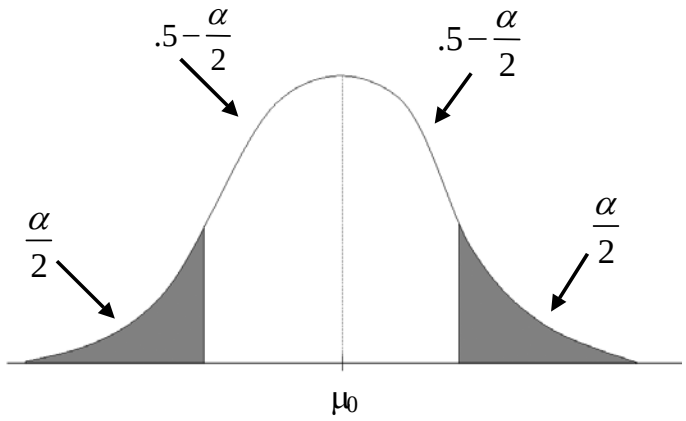
Pr(|Z|>23.28)< .0001

So, it is <u>very</u> unlikely that you'd get this sample mean if in the Washington state population female heads of households made only poverty level earnings, on average.

Classical hypothesis test- sets decision rule for accepting or rejecting null hypothesis.

- you set significance level cut-off of α (e.g., 5% or 1%)
- uses t or Z statistic cut-off for that significance level
- will reject null hypothesis if null is not within CI at same confidence level

Rejection Regions (In Grey)

2-tailed hypothesis test Is $\|z\|>z_{\alpha/2}$? Then reject the null hypothesis.	

P-value -- the probability that you'd get a sample value this far from the mean or further ***IF*** the null hypothesis were true.

- a continuous measure of "strength" of evidence for null hypothesis
- larger p value means more support for H_o
- 1-sided p values give value for one end of the distribution; 2 sided includes both.
- Will reject H_o in classical hypothesis test if $p < \alpha$

THE large sample HYPOTHESIS TEST for mean

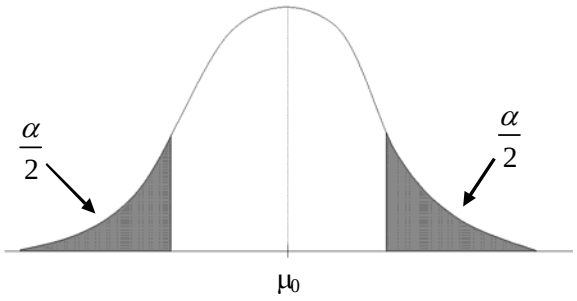
Do single female heads of households in Washington State make, on average, the poverty rate for a family of three ($16,600)?

The subsample of 1167 female-headed households who reported an average income of $36,572 with a standard deviation of $29,311.

1. Establish hypotheses: **H_0:** $\mu = \$16{,}600$ **H_a:** $\mu \neq \$16{,}600$

2. Set the decision rule:
if $|z| > z_{\alpha/2}$ then reject the null hypoth

- **pick α** $\alpha = .05$
- **find $z_{\alpha/2}$** $z_{\alpha/2} = 1.96$ (two-tailed test)
- **Draw a picture**



3. Find test statistic (like z-score)

$$z = \frac{\bar{x} - \mu_o}{SE} = \frac{36{,}572 - 16{,}600}{29{,}311/\sqrt{1167}} = \frac{19{,}972}{858} = 23.28$$

4. Compare the test statistic to critical value.

$|23.28| > 1.96$

Since $|z| > z_{\alpha/2}$ we can reject the null hypothesis at a 5% level. Could we reject the null hypothesis at the 1% level? At the 10% level?

We conclude that female-headed households in Washington State have average incomes more than the poverty level for a family of three.

THE large sample HYPOTHESIS TEST for proportion

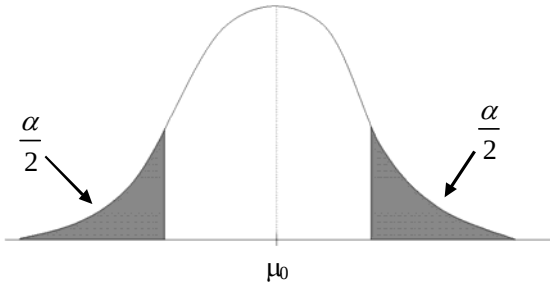
Have at least 75 percent of female heads of households in King County used the library?

We are again interested in the 1167 female-headed households, but only a subset of 573 answered the question about library use. Among these, the sample proportion who used a library was 64.9%.

1. Establish hypotheses: **H_0: p =**.75 **H_a:** p ≠ .75

2. Set the decision rule:
if $|z|>z_{\alpha/2}$ then reject the null hypothesis.

- **pick** α α=.05
- **find** $z_{\alpha/2}$ $z_{\alpha/2}$=1.96
  (two tailed test)
- **Draw a picture**



3. Find test statistic (like z-score)

$$z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{.649 - .75}{\sqrt{\frac{.75(1-.75)}{573}}} = \frac{-.101}{.0181} = -5.58$$

4. Compare the test statistic to critical value.

|-5.58| > 1.96

Since $|z| > z_{\alpha/2}$ we can reject the null hypothesis at a 5% level.

The percent of female-headed households in King County who visited the library is lower than 75%.

SPSS Output (Analyze | Compare Means | One sample t-test))

T-Test (Means)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
2005 HOUSEHOLD TOTAL INCOME	1167	36572.5870	29310.85600	858.01143

One-Sample Test

	Test Value = 16600					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
2005 HOUSEHOLD TOTAL INCOME	23.278	1166	.000	19972.58698	18289.1680	21656.0059

T-Test (Proportion)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
USED THE SERVICES OF A LIBRARY	573	.6492	.47763	.01995

One-Sample Test

	Test Value = 0.75					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
USED THE SERVICES OF A LIBRARY	-5.051	572	.000	-.10079	-.1400	-.0616

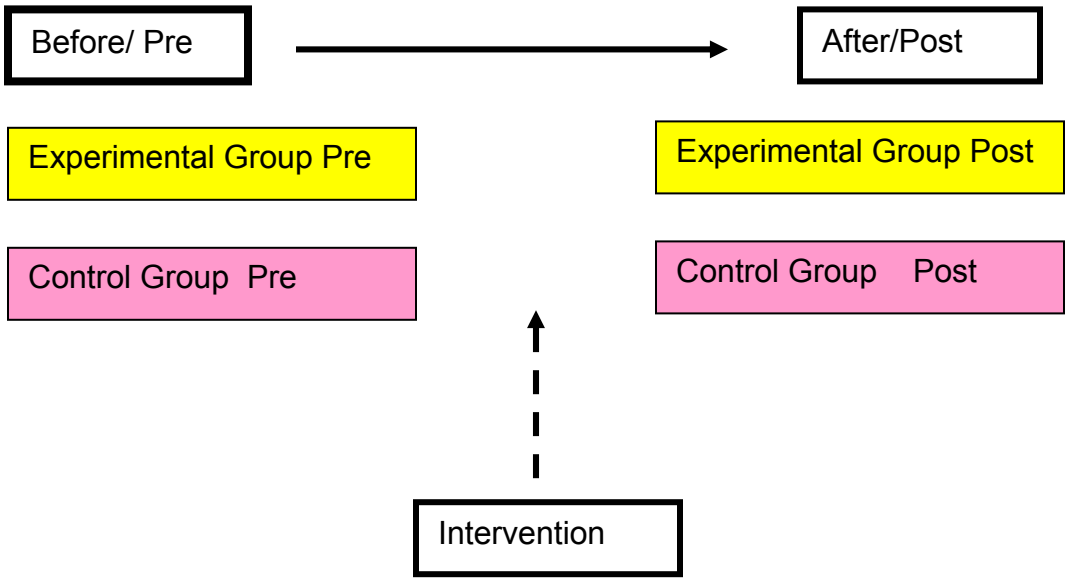
Type I and Type II Errors

- How likely is it that you'll reject a null hypothesis when it is really true?
- How likely is it that you'll accept a null hypothesis when it is really false?

Hypothesis Test Conclusion	True State of Nature	
	H_0 True	H_a True
Fail to reject H_0 (assume H_0 is true)	Correct Decision	Type II error – False negative (probability = β depends on true :)
Reject H_0 (assume H_0 is false)	Type I error -- false positive (probability = α)	Correct Decision

Research Design
We make comparisons in order to identify the sources of variation that might explain differences in important outcomes.

- Over time
- Across demographic or socioeconomic groups
- Across levels of intervention



- *What is the main research question (hypothesis) you are exploring?*
- *How are the "experimental" and "control" groups created?*
- *At what time points are you measuring outcomes?*
- *What is the "intervention" and how are you measuring it?*
- *What other factors are you accounting for and how?*
- *What other factors can you NOT account for? How does this affect your conclusions?*