

Hypothesis Testing

- Goal: Make statement(s) regarding unknown population parameter values based on sample data
- Elements of a hypothesis test:
 - **Null hypothesis** - Statement regarding the value(s) of unknown parameter(s). Typically will imply no association between explanatory and response variables in our applications (will always contain an equality)
 - **Alternative hypothesis** - Statement contradictory to the null hypothesis (will always contain an inequality)
 - **Test statistic** - Quantity based on sample data and null hypothesis used to test between null and alternative hypotheses
 - **Rejection region** - Values of the test statistic for which we reject the null in favor of the alternative hypothesis

Hypothesis Testing

Test Result –	H_0 True	H_0 False
True State H_0 True	Correct Decision	Type I Error
H_0 False	Type II Error	Correct Decision

$$\alpha = P(\text{Type I Error}) \quad \beta = P(\text{Type II Error})$$

- Goal: Keep α , β reasonably small

Example - Efficacy Test for New drug

- Drug company has new drug, wishes to compare it with current standard treatment
- Federal regulators tell company that they must demonstrate that new drug is better than current treatment to receive approval
- Firm runs clinical trial where some patients receive new drug, and others receive standard treatment
- Numeric response of therapeutic effect is obtained (higher scores are better).
- Parameter of interest: $\mu_{\text{New}} - \mu_{\text{Std}}$

Example - Efficacy Test for New drug

- **Null hypothesis** - New drug is no better than standard trt

$$H_0 : \mu_{New} - \mu_{Std} \leq 0 \quad (\mu_{New} - \mu_{Std} = 0)$$

- **Alternative hypothesis** - New drug is better than standard trt

$$H_A : \mu_{New} - \mu_{Std} > 0$$

- **Experimental (Sample) data:**

\bar{y}_{New}	\bar{y}_{Std}
s_{New}	s_{Std}
n_{New}	n_{Std}

Sampling Distribution of Difference in Means

- In large samples, the difference in two sample means is approximately normally distributed:

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

- Under the null hypothesis, $\mu_1 - \mu_2 = 0$ and:

$$Z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

- σ_1^2 and σ_2^2 are unknown and estimated by s_1^2 and s_2^2

Example - Efficacy Test for New drug

- **Type I error** - Concluding that the new drug is better than the standard (H_A) when in fact it is no better (H_0). Ineffective drug is deemed better.
 - Traditionally $\alpha = P(\text{Type I error}) = 0.05$
- **Type II error** - Failing to conclude that the new drug is better (H_A) when in fact it is. Effective drug is deemed to be no better.
 - Traditionally a clinically important difference (Δ) is assigned and sample sizes chosen so that:
$$\beta = P(\text{Type II error} \mid \mu_1 - \mu_2 = \Delta) \leq .20$$

Elements of a Hypothesis Test

- **Test Statistic** - Difference between the **Sample means**, scaled to number of standard deviations (standard errors) from the null difference of 0 for the **Population means**:

$$T.S.: z_{obs} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- **Rejection Region** - Set of values of the test statistic that are consistent with H_A , such that the probability it falls in this region when H_0 is true is α (we will always set $\alpha=0.05$)

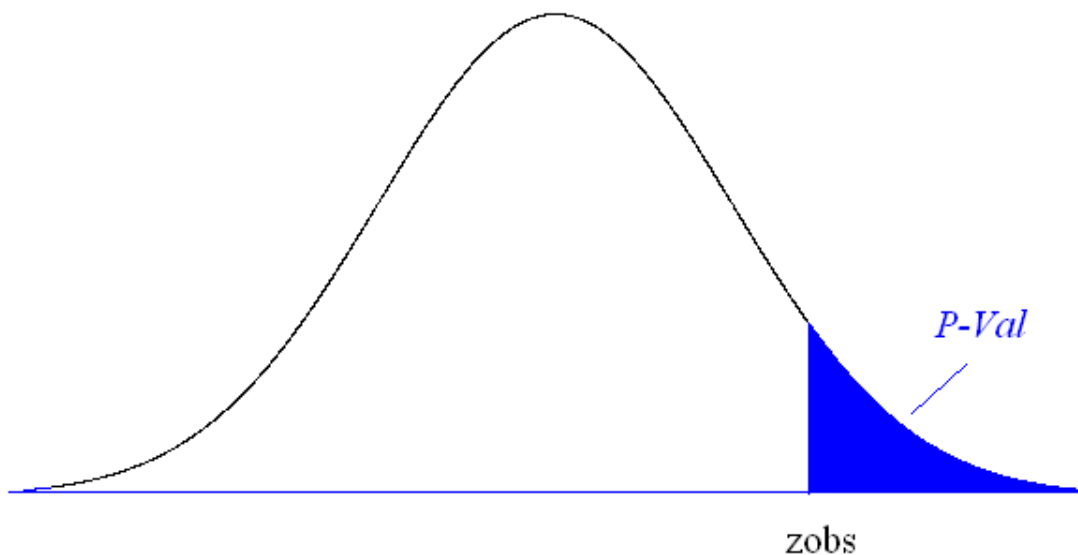
$$R.R.: z_{obs} \geq z_{\alpha} \quad \alpha = 0.05 \Rightarrow z_{\alpha} = 1.645$$

P-value (aka Observed Significance Level)

- ***P*-value** - Measure of the strength of evidence the sample data provides against the null hypothesis:

$P(\text{Evidence This strong or stronger against } H_0 \mid H_0 \text{ is true})$

$$P\text{-val} : p = P(Z \geq z_{obs})$$



Large-Sample Test $H_0: \mu_1 - \mu_2 = 0$ vs $H_A: \mu_1 - \mu_2 > 0$

- $H_0: \mu_1 - \mu_2 = 0$ (No difference in population means)
- $H_A: \mu_1 - \mu_2 > 0$ (Population Mean 1 > Pop Mean 2)

- $T.S.: z_{obs} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

- $R.R.: z_{obs} \geq z_\alpha$

- $P\text{-value}: P(Z \geq z_{obs})$

- **Conclusion** - Reject H_0 if test statistic falls in rejection region, or equivalently the P -value is $\leq \alpha$

Example - Botox for Cervical Dystonia

- **Patients** - Individuals suffering from cervical dystonia
- **Response** - Tsui score of severity of cervical dystonia (higher scores are more severe) at week 8 of Tx
- **Research (alternative) hypothesis** - Botox A decreases mean Tsui score more than placebo
- **Groups** - Placebo (Group 1) and Botox A (Group 2)
- **Experimental (Sample) Results:**

$$\bar{y}_1 = 10.1 \quad s_1 = 3.6 \quad n_1 = 33$$

$$\bar{y}_2 = 7.7 \quad s_2 = 3.4 \quad n_2 = 35$$

Example - Botox for Cervical Dystonia

Test whether Botox A produces lower mean Tsui scores than placebo ($\alpha = 0.05$)

- $H_0 : \mu_1 - \mu_2 = 0$
- $H_A : \mu_1 - \mu_2 > 0$
- $T.S. : z_{obs} = \frac{10.1 - 7.7}{\sqrt{\frac{(3.6)^2}{33} + \frac{(3.4)^2}{35}}} = \frac{2.4}{0.85} = 2.82$
- $R.R. : z_{obs} \geq z_\alpha = z_{.05} = 1.645$
- $P\text{-val} : P(Z \geq 2.82) = .0024$

Conclusion: Botox A produces lower mean Tsui scores than placebo (since $2.82 > 1.645$ and $P\text{-value} < 0.05$)

2-Sided Tests

- Many studies don't assume a direction wrt the difference $\mu_1 - \mu_2$
- $H_0: \mu_1 - \mu_2 = 0$ $H_A: \mu_1 - \mu_2 \neq 0$
- Test statistic is the same as before
- Decision Rule:
 - Conclude $\mu_1 - \mu_2 > 0$ if $z_{\text{obs}} \geq z_{\alpha/2}$ ($\alpha=0.05 \Rightarrow z_{\alpha/2}=1.96$)
 - Conclude $\mu_1 - \mu_2 < 0$ if $z_{\text{obs}} \leq -z_{\alpha/2}$ ($\alpha=0.05 \Rightarrow -z_{\alpha/2} = -1.96$)
 - Do not reject $\mu_1 - \mu_2 = 0$ if $-z_{\alpha/2} \leq z_{\text{obs}} \leq z_{\alpha/2}$
- P-value: $2P(Z \geq |z_{\text{obs}}|)$

Power of a Test

- **Power** - Probability a test rejects H_0 (depends on $\mu_1 - \mu_2$)
 - H_0 True: Power = P(Type I error) = α
 - H_0 False: Power = 1 - P(Type II error) = $1 - \beta$

- **Example:**

- $H_0: \mu_1 - \mu_2 = 0$ $H_A: \mu_1 - \mu_2 > 0$

- $\sigma_1^2 = \sigma_2^2 = 25$ $n_1 = n_2 = 25$

- Decision Rule: Reject H_0 (at $\alpha=0.05$ significance level) if:

$$z_{obs} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{2}} \geq 1.645 \quad \Rightarrow \quad \bar{y}_1 - \bar{y}_2 \geq 2.326$$

Power of a Test

- Now suppose in reality that $\mu_1 - \mu_2 = 3.0$ (H_A is true)
- Power now refers to the probability we (correctly) reject the null hypothesis. Note that the sampling distribution of the difference in sample means is approximately normal, with mean 3.0 and standard deviation (standard error) 1.414.
- Decision Rule (from last slide): Conclude population means differ if the sample mean for group 1 is at least 2.326 higher than the sample mean for group 2
- Power for this case can be computed as:

$$P(\bar{Y}_1 - \bar{Y}_2 \geq 2.326) \quad \bar{Y}_1 - \bar{Y}_2 \sim N(3, \sqrt{2.0} = 1.414)$$

Power of a Test

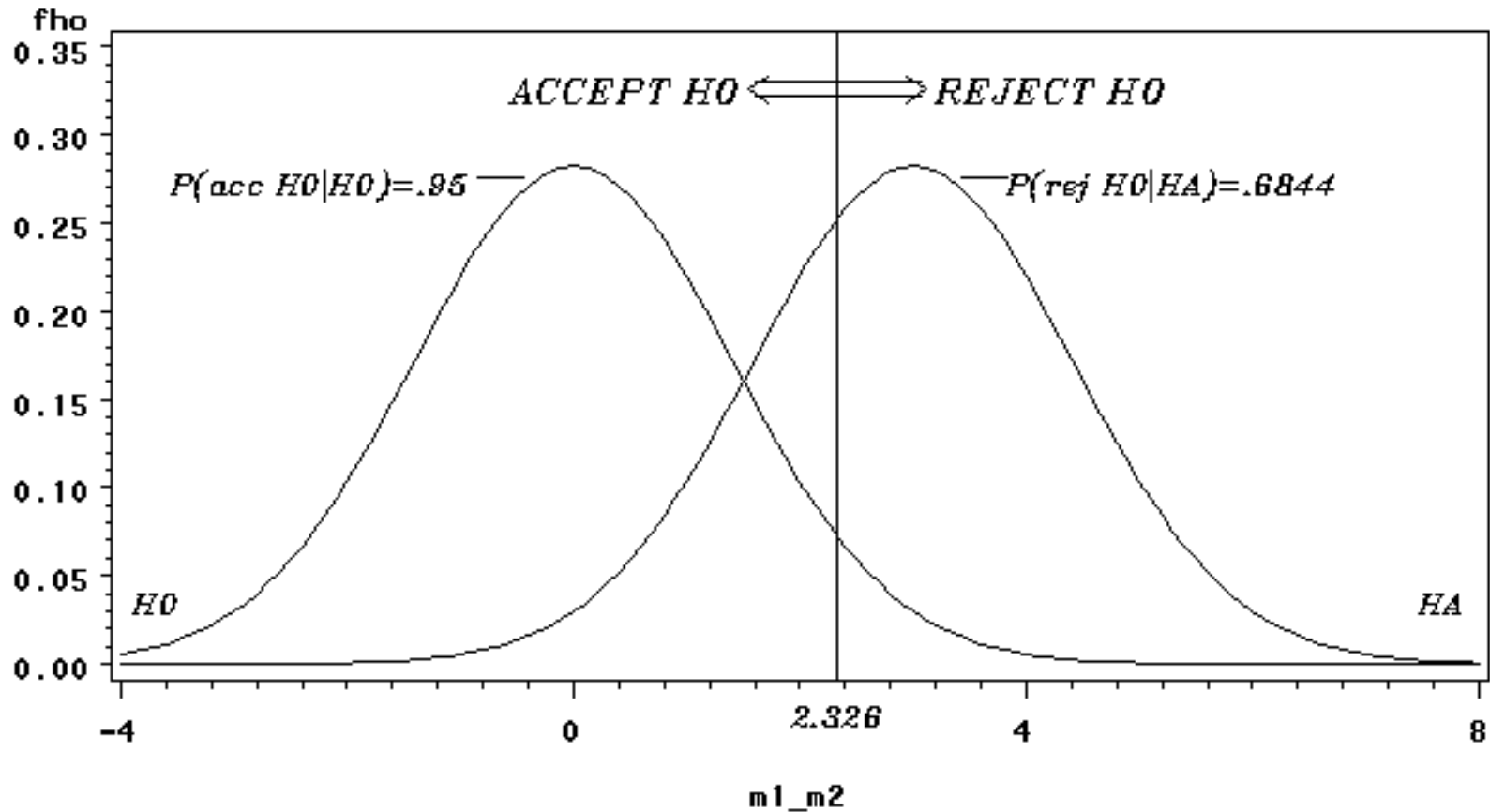
$$\text{Power} = P(\bar{Y}_1 - \bar{Y}_2 \geq 2.326) = P(Z \geq \frac{2.326 - 3}{1.41} = -0.48) = .6844$$

- All else being equal:
 - As sample sizes increase, power increases
 - As population variances decrease, power increases
 - As the true mean difference increases, power increases

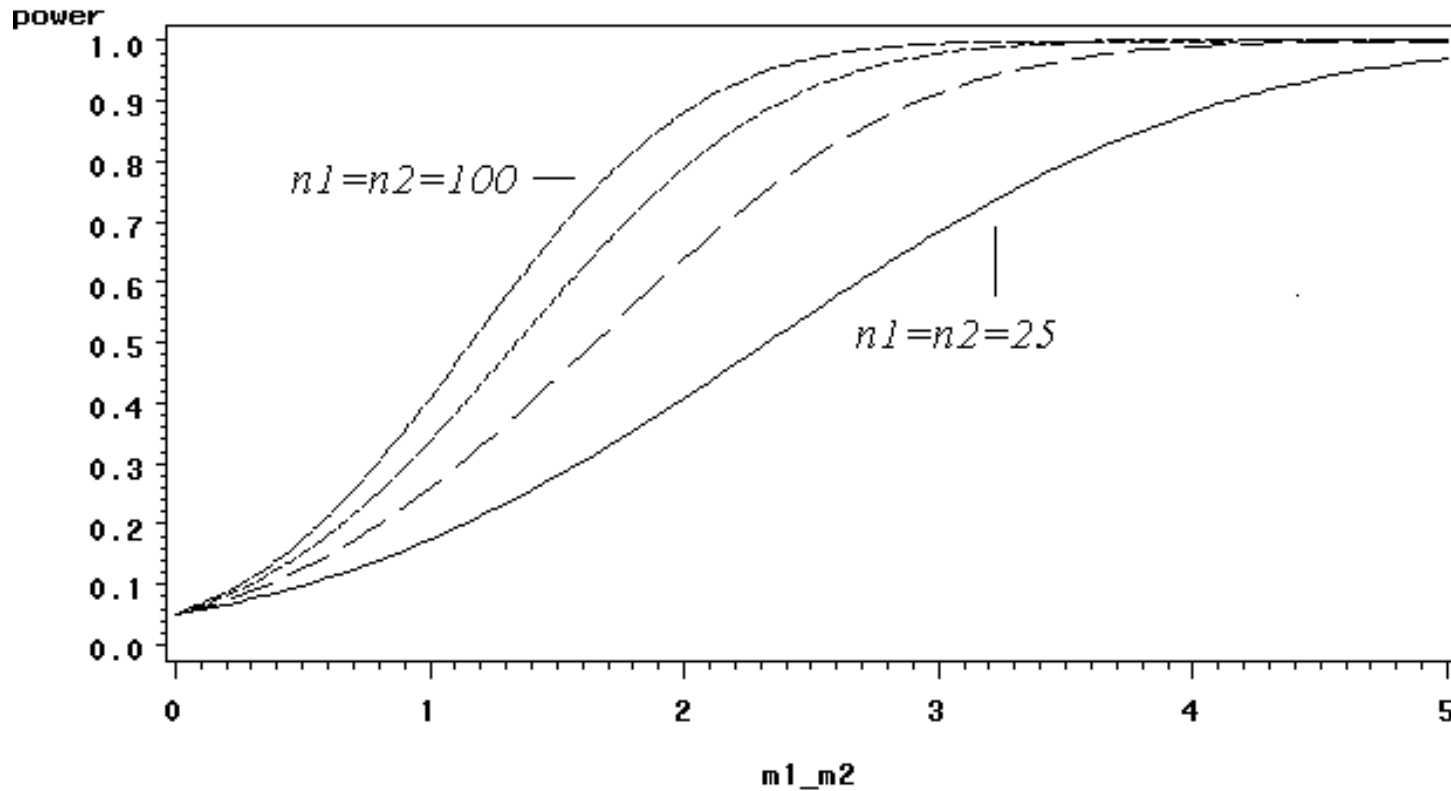
Power of a Test

Distribution (H_0)

Distribution (H_A)



Power of a Test



Power Curves for group sample sizes of 25,50,75,100 and varying true values $\mu_1 - \mu_2$ with $\sigma_1 = \sigma_2 = 5$.

- For given $\mu_1 - \mu_2$, power increases with sample size
- For given sample size, power increases with $\mu_1 - \mu_2$

Sample Size Calculations for Fixed Power

- **Goal** - Choose sample sizes to have a favorable chance of detecting a *clinically meaning difference*
- **Step 1** - Define an important difference in means:
 - **Case 1:** σ approximated from prior experience or pilot study - difference can be stated in units of the data
 - **Case 2:** σ unknown - difference must be stated in units of standard deviations of the data

$$\delta = \frac{\mu_1 - \mu_2}{\sigma}$$

- **Step 2** - Choose the desired power to detect the the clinically meaningful difference ($1-\beta$, typically at least .80). For 2-sided test:

$$n_1 = n_2 = \frac{2(z_{\alpha/2} + z_{\beta})^2}{\delta^2}$$

Example - Rosiglitazone for HIV-1 Lipoatrophy

- **Trts** - Rosiglitazone vs Placebo
- **Response** - Change in Limb fat mass
- **Clinically Meaningful Difference** - 0.5 (std dev's)
- **Desired Power** - $1-\beta = 0.80$
- **Significance Level** - $\alpha = 0.05$

$$z_{\alpha/2} = 1.96 \quad z_{\beta} = z_{.20} = .84$$

$$n_1 = n_2 = \frac{2(1.96 + 0.84)^2}{(0.5)^2} = 63$$

Confidence Intervals

- Normally Distributed data - approximately 95% of individual measurements lie within 2 standard deviations of the mean
- Difference between 2 sample means is approximately normally distributed in large samples (regardless of shape of distribution of individual measurements):

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

- Thus, we can expect (with 95% confidence) that our sample mean difference lies within 2 standard errors of the true difference

$(1-\alpha)100\%$ Confidence Interval for $\mu_1-\mu_2$

- Large sample Confidence Interval for $\mu_1-\mu_2$:

$$\left(\bar{y}_1 - \bar{y}_2\right) \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- Standard level of confidence is 95% ($z_{.025} = 1.96 \approx 2$)
- $(1-\alpha)100\%$ CI's and 2-sided tests reach the same conclusions regarding whether $\mu_1-\mu_2=0$

Example - Viagra for ED

- Comparison of Viagra (Group 1) and Placebo (Group 2) for ED
- Data pooled from 6 double-blind trials
- Subjects - White males
- Response - Percent of successful intercourse attempts in past 4 weeks (Each subject reports his own percentage)

$$\bar{y}_1 = 63.2 \quad s_1 = 41.3 \quad n_1 = 264$$

$$\bar{y}_2 = 23.5 \quad s_2 = 42.3 \quad n_2 = 240$$

95% CI for $\mu_1 - \mu_2$:

$$(63.2 - 23.5) \pm 1.96 \sqrt{\frac{(41.3)^2}{264} + \frac{(42.3)^2}{240}} \equiv 39.7 \pm 7.3 \equiv (32.4, 47.0)$$