Markov's & Chebyshev's Inequalities

- **Markov's inequality**: (Markov was a student of Chebyshev)

$$\text{If } Y \geq 0 \;\&\; d > 0 \;\Rightarrow\; P(Y \geq d) \leq \frac{E(Y)}{d}$$

$$\text{Since, if } X = \begin{cases} d, & \text{if } Y \geq d \\ 0, & \text{otherwise} \end{cases}, \quad \text{Note } Y \geq 0 \Rightarrow X \geq 0$$

$$\text{Then: } \quad E(Y) \geq E(X) \geq d \times P\{Y \geq d\}$$

$$\text{Let } Y = |X - E(X)|^2 \text{ and } d = k^2 \text{ with } k > 0 \Rightarrow$$

$$P(Y \geq d) = P(|X - E(X)|^2 \geq k^2) \leq \frac{E(|X - E(X)|^2)}{k^2} \Rightarrow$$

$$P(|X - E(X)| \geq k) \leq \frac{Var(X)}{k^2} = \frac{\sigma^2}{k^2} \Rightarrow P(|X - E(X)| \geq k \times \sigma) \leq \frac{1}{k^2}$$

Let $k' = k/\sigma \rightarrow k = k'\sigma$

Chebyshev's Theorem

- Applies to all distributions, where mean exists ($\sigma < \mu < \infty$)

Number of Standard Deviations	Distance from the Mean	Minimum Proportion of Values Falling Within Distance
K = 2	$\mu \pm 2\sigma$	$1-1/2^2 = 0.75$
K = 3	$\mu \pm 3\sigma$	$1-1/3^2 = 0.89$
K = 4	$\mu \pm 4\sigma$	$1-1/4^2 = 0.94$

Coefficient of Variation

- Ratio of the standard deviation to the mean, expressed as a percentage
- Measurement of relative dispersion

$$C.V. = \frac{\sigma}{\mu}(100)$$

Coefficient of Variation – an example

$$\mu_1 = 29 \qquad \sigma_1 = 4.6$$

$$C.V._1 = \frac{\sigma_1}{\mu_1}(100) = \frac{4.6}{29}(100) = 15.86$$

$$\mu_2 = 84 \qquad \sigma_2 = 10$$

$$C.V._2 = \frac{\sigma_2}{\mu_2}(100) = \frac{10}{84}(100) = 11.90$$

Outline

Probability Theory
- Axioms
- Basic Principles for probability modeling and computation
- Law of Total Probability & Bayesian Theorem
- Data Summaries and EDA
- Distributions (http://www.socr.ucla.edu/htmls/SOCR_Distributions.html)
- Experiments & Demos (http://www.socr.ucla.edu/htmls/SOCR_Experiments.html)

Statistical Inference
- **Parameter Estimation**
- **Hypothesis Testing & Confidence intervals**
- **Parametric vs. Non-parametric inference (http://www.socr.ucla.edu/htmls/SOCR_Analyses.html)**
- **CLT**

Linear modeling
- Simple linear regression, Multiple linear regression
- ANOVA & GLM

Parameters, Estimators, Estimates …

- A parameter is a characteristic of process, population or distribution
  - E.g., mean, 1st quartile, SD, min, max, range, skewness, 97th percentile, etc.
- An estimator is an abstract **rule** for calculating a quantity (or parameter) from sample data.
- An estimate is the value obtained when real data are plugged-in the estimator rule.

Parameters, Estimators, Estimates …

- E.g., We are interested in the population mean response time (parameter) of a cognitive experiment. The sample-average formula represents **an estimator** we can use, where as the (value of the) sample average for a particular dataset is the estimate (for the *mean* parameter).

$$\text{parameter} = \mu_Y; \qquad \text{estimator} = \bar{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k$$

$$\text{Data}: \; Y = \{0.1896,\ 0.1913,\ 0.1900\}$$

$$\text{estimate} = \bar{y} = \tfrac{1}{3}(0.1896 + 0.1913 + 0.1900)$$

$\bar{y} = 0.1903$. How about $\bar{y} = \tfrac{2}{3}(0.1896 + 0.1913 + 0.1900)$

Parameter (Point) Estimation

- Two Ways of Proposing Point Estimators
- **Method of Moments** (MOMs):
  - Set your k parameters equal to your first k moments.
  - Solve. (e.g., Binomial, Exponential and Normal)
- **Method of Maximum Likelihood** (MLEs):
- 1. Write out likelihood for sample of size n.
- 2. Take natural log of the likelihood.
- 3. Take partial derivatives with respect to your k parameters.
- 4. Take second derivatives to check that a maximum exists(f''>0).
- 5. Set 1st derivatives equal to zero and solve for MLEs. e.g., Binomial, Exponential and Normal

Parameter (Point) Estimation

- Suppose we flip a coin n=8 times and observe {T,H,T,H,H,T,H,H}. Estimate the value p = P(H).
- **Method of Moments Estimate p^**:
  - Set your k parameters equal to your first k moments.
- Let X = {# H's} ➔ np=8p=E(X)= Sample#H's = 5 ➔ p^=5/8.
- **Method of Maximum Likelihood Estimate p^**:
- 1. $f(x \mid p) = \binom{8}{5} p^5 (1-p)^3$ likelihood function.
- 2. $\ln\left(\binom{8}{5} p^5 (1-p)^3\right) = \ln\left(\binom{8}{5}\right) + 5\times\ln(p) + 3\times\ln(1-p)$
- 3. $\dfrac{d\left(\ln\left(\binom{8}{5}\right) + 5\times\ln(p) + 3\times\ln(1-p)\right)}{d\,p} = \dfrac{5}{p} - \dfrac{3}{1-p} = 0$

$5(1-p) - 3p = 0 \Rightarrow p = \tfrac{5}{8}$

Example – Maximum Likelihood Estimate

- Let $\{X_1, \ldots, X_n\} = \{0.5, 0.3, 0.6, 0.1, 0.2\}$, weights, be IID N(μ, 1) ➔ f(x;μ). **Joint density** is $f(x_1,\ldots,x_n; \mu) = f(x_1;\mu)\times\ldots\times f(x_n;\mu)$.
- The likelihood function $L(p) = f(X_1,\ldots,X_n; p)$

$$L(\mu) = \lambda(x_1,\ldots,x_n) = e^{-\frac{(0.5-\mu)^2+(0.3-\mu)^2+(0.6-\mu)^2+(0.1-\mu)^2+(0.2-\mu)^2}{2}}$$

$$\ln(L) = (-1/2)\left[(0.5-\mu)^2 + (0.3-\mu)^2 + (0.6-\mu)^2 + (0.1-\mu)^2 + (0.2-\mu)^2\right]$$

$$0 = \frac{d\ln(L)}{d\mu} = (0.5-\mu) + (0.3-\mu) + (0.6-\mu) + (0.1-\mu) + (0.2-\mu) =$$

$$= -5\mu + 1.7 \Rightarrow \mu = 0.34 \Rightarrow \frac{d^2\ln(L)}{d\mu^2} = -5 \Rightarrow L(\mu = 0.34) = \max$$

(Log)Likelihood Function

- Suppose we have a sample $\{X_1, \ldots, X_n\}$ IID $D(\theta)$ with probability density function $p = p(X \mid \theta)$. Then the joint density $p(\{X_1, \ldots, X_n\} \mid \theta)$ *is a* function of the (unknown) parameter θ.
- Likelihood function $l(\theta \mid \{X_1, \ldots, X_n\}) = p(\{X_1,\ldots,X_n\} \mid \theta)$
- Log-likelihood $L(\theta \mid \{X_1, \ldots, X_n\}) = \text{Log}_e l(\theta \mid \{X_1, \ldots, X_n\})$
- Maximum-likelihood estimation (MLE):
- Suppose $\{X_1, \ldots, X_n\}$ IID $N(\mu, \sigma^2)$, μ is unknown. We estimate it by: $MLE(\mu) = \mu\hat{} = \text{ArgMax}_\mu L(\mu \mid (\{X_1,\ldots,X_n\})$

(Log)Likelihood Function

- Suppose $\{X_1, \ldots, X_n\}$ IID $N(\mu, \sigma^2)$, μ is unknown. We estimate it by: $MLE(\mu) = \mu\hat{} = \text{ArgMax}_\mu L(\mu \mid (\{X_1,\ldots,X_n\})$

$$MLE(\mu) = Log\left(\prod_{i=1}^{n} \frac{e^{-(x_i-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}\right) = L(\mu)$$

$$0 = L'(\hat{\mu}) = \frac{1}{(2\pi\sigma^2)^{n/2}}\left(e^{-\sum_{i=1}^{n}(x_i-\hat{\mu})^2/2\sigma^2}\right)\frac{\sum_{i=1}^{n} 2(x_i-\hat{\mu})}{2\sigma^2}$$

$$\Leftrightarrow 0 = 2\sum_{i=1}^{n}(x_i - \hat{\mu}) \Leftrightarrow \hat{\mu} = \sum_{i=1}^{n} x_i / n.$$

Similarly show that : $MLE(\sigma) = \hat{\sigma} = \sum_{i=1}^{n}(x_i-\mu)^2 / n-1$.

(Log)Likelihood Function

- Suppose $\{X_1, \ldots, X_n\}$ IID Poisson(λ), λ is unknown. Estimate λ by: $MLE(\lambda)=\lambda\hat{}=\text{ArgMax}_\lambda L(\lambda|(\{X_1,\ldots,X_n\})$

$$MLE(\lambda) = Log\left(\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{(x_i)!}\right) = L(\lambda)$$

$$0 = L'(\hat{\lambda}) = \frac{\partial}{\partial\lambda} Log\left(\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}(x_i)!}\right) =$$

$$= \frac{\partial}{\partial\lambda}\left(-n\lambda + Log(\lambda)\sum_{i=1}^{n} x_i\right) = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i \Leftrightarrow \hat{\lambda} = \sum_{i=1}^{n} x_i / n$$

Hypotheses

Guiding principles

We cannot **rule in** a hypothesized value for a parameter, we *can only* determine whether there is evidence *to* **rule out** a hypothesized value.

The ***null hypothesis*** tested is typically a skeptical reaction to a ***research hypothesis***

Hypothesis Testing – the Likelihood Ratio Principle

- Let $\{X_1, \ldots, X_n\}$ be a random sample from a density f(x; p), where p is some population parameter. Then the **joint density** is $f(x_1,\ldots, x_n; p) = f(x_1; p)\times\ldots\times f(x_n; p)$.
- The likelihood function $L(p) = f(X_1,\ldots, X_n; p)$
- Testing: H_o: p is in Ω vs. H_a: p is in Ω_a, where $\Omega \cap \Omega_a = 0$
  - Find max of L(p) in Ω.
  - Find max of L(p) in Ω_a.
  - Find likelihood ratio $\lambda(x_1,\ldots,x_n) = \dfrac{\max_{p\in\Omega} L(p)}{\max_{p\in\Omega_a} L(p)}$
  - Reject H_o if likelihood-ratio statistics λ is small ($\lambda<k$)

Hypothesis Testing – the Likelihood Ratio Principle Example

- Let $\{X_1, \ldots, X_n\}=\{0.5, 0.3, 0.6, 0.1, 0.2\}$ be IID N(μ, 1) → f(x;μ). The **joint density** is $f(x_1,\ldots,x_n; \mu)=f(x_1;\mu)\times\ldots\times f(x_n;\mu)$.
- The likelihood function $L(p) = f(X_1,\ldots,X_n; p)$
- Testing: H_o: $\mu>0$ is in Ω vs H_a: $\mu<=0$.
- Reject H_o if likelihood-ratio statistics λ is small ($\lambda<k$)

$f(x) = \dfrac{e^{-(x-\mu)^2/2}}{2}$

ln(numer) = quadratic in μ!
ln(deno) = quadratic in μ!
Maximize both → find ratio

$$\lambda_0 = \lambda_{(x_1,\ldots,x_n)} = \frac{\max_{p\in\Omega} L(p)}{\max_{p\in\Omega_a} L(p)} =$$

$$\frac{\max_{\mu>0}\left(e^{-\frac{(0.5-\mu)^2+(0.3-\mu)^2+(0.6-\mu)^2+(0.1-\mu)^2+(0.2-\mu)^2}{2}}\right)}{\max_{\mu\le 0}\left(e^{-\frac{(0.5-\mu)^2+(0.3-\mu)^2+(0.6-\mu)^2+(0.1-\mu)^2+(0.2-\mu)^2}{2}}\right)}$$

Let P(Type I) = α
$t_o \sim 1/\lambda_o \sim t_{\alpha,df=4}$
→ one-sample T-test

Hypothesis Testing – the Likelihood Ratio Principle Example

- Testing: H_o: $\mu>0$ is in Ω vs H_a: $\mu<=0$. Reject H_o if likelihood-ratio statistics λ is small ($\lambda<k$)

$$\lambda_0 = \lambda_{(x_1,\ldots,x_n)} = \frac{\max_{p\in\Omega} L(p)}{\max_{p\in\Omega_a} L(p)} =$$

$$\frac{\max_{\mu>0}\left(e^{-\frac{(0.5-\mu)^2+(0.3-\mu)^2+(0.6-\mu)^2+(0.1-\mu)^2+(0.2-\mu)^2}{2}}\right)}{\max_{\mu\le 0}\left(e^{-\frac{(0.5-\mu)^2+(0.3-\mu)^2+(0.6-\mu)^2+(0.1-\mu)^2+(0.2-\mu)^2}{2}}\right)} =$$

$$\frac{e^{-\frac{(0.5-\mu)^2+(0.3-\mu)^2+(0.6-\mu)^2+(0.1-\mu)^2+(0.2-\mu)^2}{2}}\Big/_{\mu=0.34}}{e^{-\frac{(0.5-\mu)^2+(0.3-\mu)^2+(0.6-\mu)^2+(0.1-\mu)^2+(0.2-\mu)^2}{2}}\Big/_{\mu=0}} = \frac{e^{-0.086}}{e^{-0.375}} = e^{-0.269} = 0.68$$

Let P(Type I) = α
$t_o \sim 1/\lambda_o \sim t_{\alpha,df=4}$
→ one-sample T-test

Inference and Hypothesis Testing

1. Identify your design & appropriate statistical technique
   http://www.socr.ucla.edu/htmls/SOCR_ChoiceOfStatisticalTest.html
2. Validate your Data/Model Assumptions
3. Calculate a Test Statistic (Example: z_o)
4. Specify a Rejection Region (Example: $|z|>z_{\alpha/2}$)
5. Inference: The null hypothesis is rejected iff the computed value for the statistic falls in the rejection region

Type I and Type II Errors

$$\alpha = \Pr\{\text{Reject } H_0 | H_0 \text{ is true}\}$$
$$\beta = \Pr\{\text{Fail to Reject } H_0 | H_0 \text{ is False}\}$$

- The value of α is specified by the experimenter
- The value of β is a function of **α, n, and δ** (the difference between the null hypothesized mean and the true mean). For a two sided hypothesis test of a normally distributed population

$$\beta = \Phi\left(Z_{\frac{\alpha}{2}} + \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-Z_{\frac{\alpha}{2}} + \frac{\delta\sqrt{n}}{\sigma}\right)$$

- It is not true that $\alpha = 1 - \beta$ (RHS=this is the test power!)

Type I, Type II Errors & Power of Tests

- Suppose the true MMSE score for AD subjects is ~ N(23, 1^2).
- A new cognitive test is proposed, and it's assumed that its values are ~ N(25, 1^2). A sample of 10 AD subjects take the new test.
- Hypotheses are: H_o: μ_{test}=25 vs. H_a: μ_{test}<25 (one-sided, more power)
- α = P(false-positive, Type I, error) = 0.05.
- Critical Value for α is Z_{score}= -1.64. Thus, $X^{avg}_{critical} = Z_{critical} * \sigma + \mu$
- $X^{avg}_{critical}$ = 25-1.64=23.4, And our conclusion, from $\{X_1, \ldots, X_{10}\}$ which yields X^{avg} will be **reject** H_o, if X^{avg} < 23.4.
- β=P(fail to reject H_o | H_o is false)=P(X^{avg}>=23.4 | X^{avg} ~N(23,1^2/10))
- Note: X^{avg} ~N(23,1^2/10)), when it's given that X ~N(23,1^2))
- Standardize: Z = (23.4 - 23)/(1/10) = 4.0

Type I, Type II Errors & Power of Tests

- Suppose the true MMSE score for AD subjects is ~ N(23, 1^2).
- A new cognitive test is proposed, and it's assumed that its values are ~ N(25, 1^2). A sample of 10 AD subjects take the new test.
- β=P(fail to reject H_o | H_o is false)=P(X^{avg}>=23.4 | X^{avg} ~N(23,1^2/10))
- Note: X^{avg} ~N(23,1^2/10)) when it's given that X ~N(23,1^2))
- Standardize: Z = (23.4 - 23)/(1/10) = 4.0.
- β=P(fail to reject H_o | H_o is false)=P(Z>4.0) = 0.00003
- ➔ Power (New Test) = 1 - 0.00003 = 0.99997
- **How does Power(Test) depend on:**
  - Sample size, n=10: n-increase ➔ power increase
  - Size-of-studied-effect: effect-size increase ➔ power increase
  - Type of Alternative hypothesis: 1-sized tests are more powerful

∃different β for each different α, true mean μ, alternative H_a

Another Example -Type I and II Errors & Power

- About 75% of all 80 year old humans are free of amyloid plaques and tangles, markers of AD. A new AD vaccine is proposed that is supposed to increase this proportion. Let p be the new proportion of subjects with no AD characteristics following vaccination. H_o: p=0.75, H_1: p>0.75.
- X = number of AD tests with no pathology findings in n=20 80-y/o vaccinated subjects. Under H_o we expect to get about **n*p=15 no AD results**. Suppose we'd invest in the new vaccine if we get >= 18 no AD tests ➔ rejection region R={18, 19, 20}.
- Find α and β. How powerful is this test?

Another Example -Type I and Type II Errors

- H_o: p=0.75, H_1: p>0.75. X = number of test with no AD findings in n=20 experiments.
- X~Binomial(20, 0.75). Rejection region R={18, 19, 20}.
- Find α =P(Type I) = P(X>=18 when X~Binomial(20, 0.75)).
- Use SOCR resource ➔ **α =1-0.91= 0.09** How does Power(Test) depend on n, effect-size?
- Find β(p=0.85) =P (Type II) =
  - P(fail to reject H_o | X~Binomial(20, 0.85))=P(X<18 | X~Binomial(20, 0.85))
  - Use SOCR resource ➔ β =0.595➔Power of test = 1- β =0.405
- Find β(p=0.95) =P (Type II) =
  - P(fail to reject H_o | X~Binomial(20, 0.95))=P(X<18 | X~Binomial(20,0.95))
  - Use SOCR resource ➔ β =0.076 ➔Power of test = 1- β = 0.924

A 95% confidence interval

- A type of interval that contains the true value of a parameter for 95% of samples taken is called a ***95% confidence interval*** for that parameter, the ends of the CI are called *confidence limits*.
- (For the situations we deal with) a confidence interval (CI) for the true value of a **parameter** is given by

estimate $\pm$ ***t standard errors (SE)***

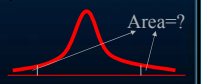
Value of the Multiplier, *t*, for a 95% CI

df :	7	8	9	10	11	12	13	14	15	16	17
t :	2.365	2.306	2.262	2.228	2.201	2.179	2.160	2.145	2.131	2.120	2.110
df :	18	19	20	25	30	35	40	45	50	60	∞
t :	2.101	2.093	2.086	2.060	2.042	2.030	2.021	2.014	2.009	2.000	1.960

(General) Confidence Interval (CI)

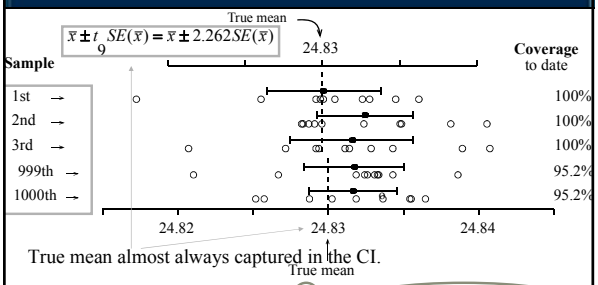
- A **level L confidence interval** for a parameter (θ), is an interval (θ_1^ , θ_2^), where θ_1^ & θ_2^, are estimators of θ, such that $P(\theta_1\text{^} < \theta < \theta_2\text{^}) = L$.
- E.g., **C+E model**: $Y = \mu+\varepsilon$. Where $\varepsilon \sim N(0, \sigma^2)$, then by CLT we have $Y\_bar \sim N(\mu, \sigma^2/n)$

  → $n^{1/2}(Y\_bar - \mu)/\sigma \sim N(0, \sigma^2)$.



- $L = P\left(z_{(1-L)/2} < n^{1/2}(Y\_bar - \mu)/\sigma < z_{(1+L)/2} \right)$, where z_q is the q^{th} quartile.
- E.g., $0.95 = P\left(z_{0.025} < n^{1/2}(Y\_bar - \mu)/\sigma < z_{0.975} \right)$.

- CI are constructed using the sample $\bar{x}$ and s=SE. But different samples yield different estimates and → diff. CI's?!?
- Below is a computer simulation showing how the process of taking samples effects the estimates and the CI's.



Confidence Interval

Confidence Interval (CI) Experiment
http://socr.ucla.edu/htmls/SOCR_Experiments.html

CI Activity:
http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_ConfidenceIntervals

Comparing two means for independent samples

Suppose we have 2 samples/means/distributions as follows: $\{\bar{x}_1, N(\mu_1, \sigma_1)\}$ and $\{\bar{x}_2, N(\mu_2, \sigma_2)\}$. We've seen before that to make inference about $\mu_1 - \mu_2$ we can use a T-test for H_0: $\mu_1 - \mu_2 = 0$ with

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{SE(\bar{x}_1 - \bar{x}_2)}$$

And CI($\mu_1 - \mu_2$) = $\bar{x}_1 - \bar{x}_2 \pm t \times SE(\bar{x}_1 - \bar{x}_2)$

If the 2 samples are independent we use the SE formula

$$SE = \sqrt{s_1^2/n_1 + s_2^2/n_2} \quad \text{with} \quad df = Min(n_1 - 1; n_2 - 1)$$

This gives a conservative approach for hand calculation of an approximation to the what is known as the Welch procedure, which has a complicated exact formula.

Means for independent samples - equal or unequal variances?

Pooled T-test is used for samples with assumed equal variances. Under data Normal assumptions and equal variances of

$$(\bar{x}_1 - \bar{x}_2 - 0)/SE(\bar{x}_1 - \bar{x}_2), \text{where}$$

$$SE = s_p\sqrt{1/n_1 + 1/n_2}; s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$$

is exactly Student's t distributed with $df = (n_1 + n_2 - 2)$

Here s_p is called the pooled estimate of the variance, since it pools info from the 2 samples to form a combined estimate of the single variance $\sigma_1^2 = \sigma_2^2 = \sigma^2$. The book recommends routine use of the Welch unequal variance method.

Single Sample: Testing/CI

- **Example**: Suppose a researcher is interested in studying the effect of aspirin in **reducing heart attacks**. He randomly recruits **500** subjects with evidence of early heart disease and has them take one aspirin daily for two years. At the end of the two years he finds that during the study only **17** subjects had a heart attack.
- Calculate a **95% confidence interval** for the true proportion of subjects with early heart disease that have a heart attack while taking aspirin daily.

Single Sample: Testing/CI

Example: Heart Attacks (cont')

First, we need to find $z_{\alpha/2}$

- because this is a 95% CI, this means that α will be 0.05 and $z_{\alpha/2}$ will be $z_{0.025}$

- in this case $z_{\alpha/2} = 1.96$

Single Sample: Testing/CI

Next, solve for $\tilde{p}$

Often rounded to $\frac{y+2}{n+4}$

$$\tilde{p} = \frac{y + 0.5\left(z^2_{\alpha/2}\right)}{n + z^2_{\alpha/2}} = \frac{y + 0.5\left(z^2_{0.025}\right)}{n + z^2_{0.025}} = \frac{y + 0.5\left(1.96^2\right)}{n + 1.96^2} = \frac{y + 1.92}{n + 3.84}$$

- that's just the formula for $\tilde{p}$, now we actually have to find $\tilde{p}$

$$\tilde{p} = \frac{17 + 1.92}{500 + 3.84} = 0.038$$

Single Sample: Testing/CI

Next, solve for $SE_{\tilde{p}}$

$$SE_{\tilde{p}} = \sqrt{\frac{(0.038)(0.962)}{500 + 3.84}} = 0.0085$$

Finally the 95% CI for p

$$\tilde{p} \pm z_{\alpha/2}\left(SE_{\tilde{p}}\right) = 0.038 \pm 1.96(0.0085)$$
$$= 0.038 \pm 0.0167 = (0.0213,\ 0.0547)$$

Single Sample: Testing/CI

What is our interpretation of this interval?

CONCLUSION: We are highly confident, at the 0.05 level (95% confidence), that the true proportion of subjects with early heart disease who have a heart attack after taking aspirin daily is between 0.0213 and 0.0547.

- Is this meaningful?

Comparison of Two Independent Samples

- Two Approaches for Comparison
  - What seems like a reasonable way to compare two groups?
- What parameter are we trying to estimate?

Comparison of Two Independent Samples

RECALL: The sampling distribution of $\bar{y}$ was centered at μ, and had a standard deviation of $\sigma/\sqrt{n}$

We'll start by describing the sampling distribution of $\bar{y}_1 - \bar{y}_2$

- Mean: $\mu_1 - \mu_2$
- Standard deviation of $\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$

What seems like appropriate estimates for these quantities?

Standard Error of $\bar{y}_1 - \bar{y}_2$

We know $\bar{y}_1 - \bar{y}_2$ estimates $\mu_1 - \mu_2$

What we need to describe next is the precision of our estimate, $SE_{(\bar{y}_1 - \bar{y}_2)}$

$$SE_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{SE_1^2 + SE_2^2}$$

Standard Error of $\bar{y}_1 - \bar{y}_2$

Example: A study is conducted to quantify the benefits of a new cholesterol lowering medication. Two groups of subjects are compared, those who took the medication twice a day for 3 years, and those who took a placebo. Assume subjects were randomly assigned to either group and that both groups data are normally distributed. Results from the study are shown below:

	Medication	Placebo
$\bar{y}$	209.8	224.3
n	10	10
s	44.3	46.2
SE	14.0	14.6

Standard Error of $\bar{y}_1 - \bar{y}_2$

Example: Cholesterol medicine (cont')

(e.g., ftp://ftp.nist.gov/pub/dataplot/other/reference/CHOLEST2.DAT)

Calculate an estimate of the true mean difference between treatment groups and this estimate's precision.

- First, denote medication as group 1 and placebo as group 2

	Medication	Placebo
$\bar{y}$	209.8	224.3
n	10	10
s	44.3	46.2
SE	14.0	14.6

$$(\bar{y}_1 - \bar{y}_2) = 209.8 - 224.3 = -14.5$$

$$SE_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{44.3^2}{10} + \frac{46.2^2}{10}} = 20.24$$

Pooled vs. Unpooled

$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ is know as an ***unpooled*** version of the standard error

- there is also a "pooled" SE

First we describe a **pooled** variance, which can be thought of as a weighted average of s_1^2 and s_2^2

$$s_{pooled}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Pooled vs. Unpooled

Then we use the pooled variance to calculate the pooled version of the standard error

$$SE_{pooled} = \sqrt{s_{pooled}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

- NOTE: If ($n_1 = n_2$) and ($s_1 = s_2$) the pooled and unpooled will give the same answer for $SE_{(\bar{y}_1 - \bar{y}_2)}$
- It is when $n_1 \neq n_2$ that we need to decide whether to use pooled or unpooled:
  - if $s_1 = s_2$ then use pooled (unpooled will give similar answer)
  - if $s_1 \neq s_2$ then use unpooled (pooled will NOT give similar answer)

Pooled vs. Unpooled

- RESULT: Because both methods are similar when $s_1 = s_2$ and $n_1 = n_2$, and the pooled version is not valid when
- Why all the torture? This will come up again in chapter 11.
- Because the ***df*** increases a great deal when we do pool the variance.

CI for μ_1 - μ_2

RECALL: We described a CI earlier as:
the estimate ± (an appropriate multiplier)x(SE)

A 100(1- α)% confidence interval for μ_1 - μ_2
(p.227)

$$(\bar{y}_1 - \bar{y}_2) \pm t(df)_{\alpha/2}\left(SE_{\bar{y}_1-\bar{y}_2}\right)$$

where df = $$\frac{\left(SE_1^2 + SE_2^2\right)^2}{\frac{SE_1^4}{(n_1-1)} + \frac{SE_2^4}{(n_2-1)}}$$

CI for μ_1 - μ_2

Example: Cholesterol medication (cont')

Calculate a 95% confidence interval for μ_1 - μ_2

We know $\bar{y}_1 - \bar{y}_2$ and $SE_{(\bar{y}_1-\bar{y}_2)}$ from the previous slides. Now we need to find the t multiplier

$$df = \frac{\left(14^2 + 14.6^2\right)^2}{\frac{14^4}{(10-1)} + \frac{14.6^4}{(10-1)}} = \frac{167411.9056}{9317.021} = 17.97 \approx 17$$

Round down to be conservative

•NOTE: Calculating that df is not really that fun, a quick rule of thumb for checking your work is:

$n_1 + n_2 - 2$

CI for μ_1 - μ_2

$$(\bar{y}_1 - \bar{y}_2) \pm t(df)_{\alpha/2}\left(SE_{\bar{y}_1-\bar{y}_2}\right)$$
$$= -14.5 \pm t(17)_{0.025}(20.24)$$
$$= -14.5 \pm 2.110(20.24)$$
$$= (-57.21,\ 28.21)$$

CONCLUSION: We are highly confident at the 0.05 level, that the true mean difference in cholesterol between the medication and placebo groups is between -57.02 and 28.02 mg/dL.

Note the change in the conclusion of the parameter that we are estimating. Still looking for the 5 basic parts of a CI conclusion (see slide 38 of lecture set 5).

CI for μ_1 - μ_2

- What's so great about this type of confidence interval?
- In the previous example our CI contained zero
  - This interval isn't telling us much because:
    - the true mean difference could be more than zero (in which case the mean of group 1 is larger than the mean of group 2)
    - or the true mean difference could be less than zero (in which case the mean of group 1 is smaller than the mean of group 2)
    - or the true mean difference could even be zero!
  - •The ZERO RULE!
  - Suppose the CI came out to be (5.2, 28.1), would this indicate a true mean difference?

Hypothesis Testing: The independent t test

- The idea of a hypothesis test is to formulate a hypothesis that nothing is going on and then to see if collected data is consistent with this hypothesis (or if the data shows something different)
  - Like innocent until proven guilty
- There are four main parts to a hypothesis test:
  - hypotheses
  - test statistic
  - p-value
  - conclusion

Hypothesis Testing: #1 The Hypotheses

- There are two hypotheses:
  - Null hypothesis (aka the "status quo" hypothesis)
    - denoted by H_o
  - Alternative hypothesis (aka the research hypothesis)
    - denoted by H_a

Hypothesis Testing: #1 The Hypotheses

- If we are comparing two group means nothing going on would imply no difference
  - the means are "the same"
  - $(\mu_1 - \mu_2) = 0$
- For the independent t-test the hypotheses are:
  - H_o: $(\mu_1 - \mu_2) = 0$
  - (no statistical difference in the population means)
  - H_a: $(\mu_1 - \mu_2) \neq 0$
  - (a statistical difference in the population means)

Hypothesis Testing: #1 The Hypotheses

Example: Cholesterol medication (cont')

Suppose we want to carry out a hypothesis test to see if the data show that there is enough evidence to support a difference in treatment means.

Find the appropriate null and alternative hypotheses.

H_o: $(\mu_1 - \mu_2) = 0$

(no statistical difference the true means of the medication and placebo groups)

H_a: $(\mu_1 - \mu_2) \neq 0$

(a statistical difference in the true means of the medication and placebo groups, medication has an effect on cholesterol)

Hypothesis Testing: #2 Test Statistic

A test statistic is calculated from the sample data

- it measures the "disagreement" between the data and the null hypothesis
  - if there is a lot of "disagreement" then we would think that the data provide evidence that the null hypothesis is false
  - if there is little to no "disagreement" then we would think that the data do not provide evidence that the null hypothesis is false

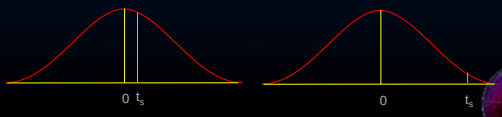
$$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{\bar{y}_1 - \bar{y}_2}}$$

subtract 0 because the null says the difference is zero

Hypothesis Testing: #2 Test Statistic

On a t distribution t_s could fall anywhere

- If the test statistic is close to 0, this shows that the data are compatible with H_o (no difference)
  - the deviation can be attributed to chance
- If the test statistic is far from 0 (in the tails, upper or lower), this shows that the data are incompatible to H_o (there is a difference)
  - deviation does not appear to be attributed to chance (ie. If H_o is true then it is unlikely that t_s would fall so far from 0)



Hypothesis Testing: #2 Test Statistic

Example: Cholesterol medication (cont')

Calculate the test statistic

$$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{\bar{y}_1 - \bar{y}_2}} = \frac{(209.8 - 224.3) - 0}{20.24} = -0.716$$

- Great, what does this mean?
  - $\bar{y}_1$ and $\bar{y}_2$ differ by about 0.72 SE's
  - this is because t_s is the measure of difference between the sample means expressed in terms of the SE of the difference

Hypothesis Testing: #2 Test Statistic

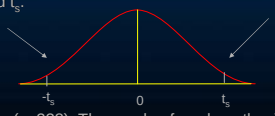
How do we use this information to decide if the data support H_o?

- Perfect agreement between the means would indicate that $t_s = 0$, but logically we expect the means do differ by at least a little bit.
  - The question is how much difference is statistically significant?
- If H_o is true, it is unlikely that t_s would fall in either of the far tails
- If H_o is false it is unlikely that t_s would fall near 0

Hypothesis Testing: #3 P-value

How far is far?

For a two tailed test (i.e. H_a: $(\mu_1 - \mu_2) \neq 0$) The p-value of the test is the area under the Student's T distribution in the double tails beyond $-t_s$ and t_s.

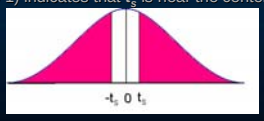


- Definition (p. 238): The p-value for a hypothesis test is the probability, computed under the condition that the null hypothesis is true, of the test statistic being at least as extreme or more extreme as the value of the test statistic that was actually obtained [from the data].

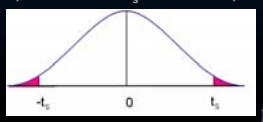
Hypothesis Testing: #3 P-value

What this means is that we can think of the p-value as a measure of compatibility between the data and H_o

- a large p-value (close to 1) indicates that t_s is near the center (data support H_o)



- a small p-value (close to 0) indicates that t_s is in the tail (data do not support H_o)



Hypothesis Testing: #3 P-value

Where do we draw the line?

- how small is small for a p-value?

The threshold value on the p-value scale is called the significance level, and is denoted by *a*

- The significance level is chosen by whomever is making the decision (BEFORE THE DATA ARE COLLECTED!)
- Common values for include 0.1, 0.05 and 0.01

Rules for making a decision:

If $p \leq a$ then reject H_o (statistical significance)

If $p > a$ then fail to reject H_o (no statistical significance)

Hypothesis Testing: #3 P-value

Example: Cholesterol medication (cont')

Find the p-value that corresponds to the results of the cholesterol lowering medication experiment

We know from the previous slides that t = -0.716 (which is close to 0)

This means that the p-value is the area under the curve beyond $\pm$ 0.716 with 18 df.

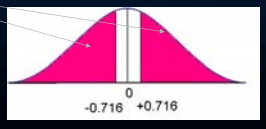
Hypothesis Testing: #3 P-value

Example: Cholesterol medication (cont')

Using SOCR we can find the area under the curve beyond $\pm$ 0.716 with 18 df to be:

p > 2(0.2) = 0.4

NOTE, when H_a is $\neq$, the p-value is the area beyond the test statistic in BOTH tails.



Hypothesis Testing: #4 Conclusion

Example: Cholesterol medication (cont')

Suppose the researchers had set α = 0.05

Our decision would be to fail to reject Ho because p > 0.4 which is > 0.05

(#4) CONCLUSION: Based on this data there is no statistically significant difference between true mean cholesterol of the medication and placebo groups (p > 0.4).

- In other words the cholesterol lowering medication does not seem to have a significant effect on cholesterol.

• Keep in mind, we are saying that we couldn't provide sufficient evidence to show that there is a significant difference between the two *population* means.

Comparing two means for independent samples

1. How sensitive is the two-sample *t*-test to non-Normality in the data? (The 2-sample T-tests and CI's are even more robust than the 1-sample tests, against non-Normality, particularly when the shapes of the 2 distributions are similar and $n_1=n_2=n$, even for small n, remember $df = n_1+n_2-2$.
3. Are there nonparametric alternatives to the *two-sample t-test*? (Wilcoxon rank-sum-test, Mann-Witney test, equivalent tests, same P-values.)
4. What difference is there between the quantities tested and estimated by the two-sample *t*-procedures and the nonparametric equivalent? (Non-parametric tests are based on ordering, not size, of the data and hence use median, not mean, for the average. The equality of 2 means is tested and CI($\mu_1 - \mu_1$).

Paired Comparisons

- An fMRI study of N subjects: The point in the time course of maximal activation in the rostral and caudal medial premotor cortex was identified, and the percentage changes in response to the ***go*** and ***no-go*** tasks from the rest state measured. Similarly the points of maximal activity during the ***go*** and ***no-go*** task were identified in the primary motor cortex. Paired t-test comparisons between the ***go*** and ***no-go*** percentage changes were performed across subjects for these regions of maximum activity.

Paired data

- We have to distinguish between independent and related samples because they require different methods of analysis.
- Paired data is an example of related data.
- With paired data, we analyze the differences
  - this converts the initial problem into a one-sample problem.
- The ***sign test*** and ***Wilcoxon rank-sum*** test are nonparametric alternatives to the and paired *t*-test, and independent *t*-test, respectively.

The Wilcoxon-Mann-Whitney

- Also known as the rank sum test
- http://www.socr.ucla.edu/Applets.dir/WilcoxonRankSumTable.html
- This hypothesis test is **also used to compare two independent samples**
  - This procedure is different from the *independent t test* because it is **valid even if the population distributions are not normal**
  - In other words, we can use this test as a fair substitute when we cannot not meet the required normality assumption of the t test

WMW is called a **distribution-free** type of test or a non-parametric test

- This test doesn't focus on a parameter like the mean, instead it examines the distributions of the two groups

The Wilcoxon-Mann-Whitney

Keep in mind that this is another hypothesis test, there are four major parts to consider

#1 The hypotheses:

- H_o: The population distributions of Y_1 and Y_2 are the same
- H_a: The population distributions of Y_1 and Y_2 are the different
  - This could also be directional: distribution of Y_1 is less than Y_2; OR distribution of Y_1 is greater than Y_2

#2 The test statistic:

- denoted by U_s
- measures the degree of separation between the two samples
  - a large value of U_s indicates that the two samples are well separated with little overlap
  - a small value of U_s indicates that the two samples are not well separated with much overlap

The Wilcoxon-Mann-Whitney

#3 The p-value:

- http://www.socr.ucla.edu/Applets.dir/WilcoxonRankSumTable.html
- Method very similar to using the t table
  - find the appropriate row and then search for a number closest to the test statistic
- don't need to worry about doubling the p-value for a two-tailed test (assuming we go to the right row header)

#4 Conclusion:

- Similar to the conclusion of an independent t test, but not linked to any parameter (for example the difference in means)

The Wilcoxon-Mann-Whitney

The Method:

- Step 1: Arrange the data in increasing order
- Step 2: Determine K_1 and K_2
  - K_1: for each observation in group 1, count the number of observations in the second group that are smaller. Use 1/2 for tied observations. K_1 is the sum of these ranks.
  - K_2: for each observation in group 2, count the number of observations in the first group that are smaller. Use 1/2 for tied observations. K_2 is the sum of these ranks.
  - CHECK: if you have done the procedure correctly $K_1 + K_2 = n_1 n_2$
- Step 3: If the test is non-directional then U_s is the larger of K_1 and K_2. If the test is directional then U_s is the K that jives with the direction of H_a (**if H_a is $Y_1 > Y_2$ then $U_s = K_1$, if H_a is $Y_1 < Y_2$ then $U_s = K_2$**)
- Step 4: Determine the critical value
  - n = larger of n_1 and n_2
  - n' = smaller of n_1 and n_2
- Step 5: Bracket the p-value

The Wilcoxon-Mann-Whitney

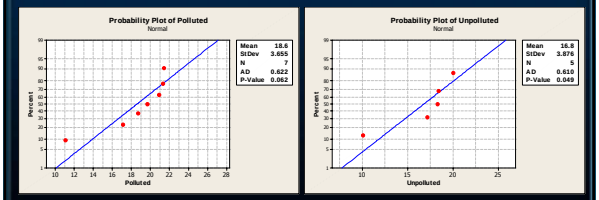
Example: The urinary fluoride concentration (ppm) was measured both for a sample of livestock grazing in an area previously exposed to fluoride pollution and also for a similar sample of livestock grazing in an unpolluted area.

Polluted	Unpolluted
21.3	10.1
18.7	18.3
21.4	17.2
17.1	18.4
11.1	20.0
20.9	
19.7	

Does the data suggest that the fluoride concentration for livestock grazing in the polluted region is larger that for the unpolluted region? Test using $\alpha = 0.01$.

The Wilcoxon-Mann-Whitney

Check Normality:



The Wilcoxon-Mann-Whitney

Conditions for the WMW:

- Data are from random samples
- Observations are independent
- Samples are independent

Remember: normality will not matter for this test

Wilcoxon-Mann-Whitney vs. Independent T-Test

Both try to answer the same question, but treat data differently.

- W-M-W uses rank ordering
  - Pro: doesn't depend on normality or population parameters
  - Con: distribution free lacks power because it doesn't use all the info in the data
- T-test uses actual Y values
  - Pro : Incorporates all of the data into calculations
  - Con : Must meet normality assumption
- neither is superior

So…

- If your data are normally distributed use the t-test
- If your data are not normal use the WMW test

The Sign Test

http://www.socr.ucla.edu/htmls/SOCR_Analyses.html

The **sign test is a non-parametric alternative of the paired t test**

We use the sign test when pairing is appropriate, but we can't meet the normality assumption required for the t test

The sign test is not very sophisticated and therefore quite easy to understand

Sign test is also based on differences

$d = Y_1 - Y_2$

The information used by the sign test from this difference is the sign of d (+ or -)

The Sign Test

#1 Hypotheses:

H_o: the distributions of the two groups is the same

H_a: the distributions of the two groups is different

or H_a: the distribution of group 1 is less than group 2

or H_a: the distribution of group 1 is greater than group 2

#2 Test Statistic B_s

The Sign Test - Method

#2 Test Statistic B_s:

1. Find the sign of the differences
2. Calculate N_+ and N_-
3. If H_a is non-directional, B_s is the larger of N_+ and N_-
   If H_a is directional, B_s is the N that jives with the direction of Ha:
   if H_a: $Y_1<Y_2$ then we expect a larger N_-,
   if H_a: $Y_1>Y_2$ then we expect a larger N_+.

NOTE: If we have a difference of zero it is not included in N_+ or N_-, therefore n_d needs to be adjusted

The Sign Test

#3 p-value:

Similar to the WMW

Use the number of pairs with "quality information"

#4 Conclusion:

Similar to the Wilcoxon-Mann-Whitney Test

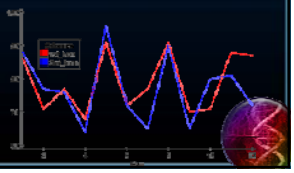
Do NOT mention any parameters!

The Sign Test

Example: 12 sets of *identical twins* are given psychological tests to determine whether the first born of the set tends to be more aggressive than the second born. Each twin is scored according to aggressiveness, a higher score indicates greater aggressiveness.

Because of the natural pairing in a set of twins these data can be considered paired.

Set	1st born	2nd born	Sign of d
1	86	88	-
2	71	77	-
3	77	76	+
4	68	64	+
5	91	96	-
6	72	72	Drop
7	77	65	+
8	91	90	+
9	70	65	+
10	71	80	-
11	88	81	+
12	87	72	+



The Sign Test (cont')

Do the data provide sufficient evidence to indicate that the first born of a set of twins is more aggressive than the second? Test using $\alpha = 0.05$.

- H_o: The aggressiveness is the same for 1st born and 2nd born twins
- H_a: The aggressiveness of the 1st born twin tends to be more than 2nd born.
- NOTE: Directional Ha (we're expecting higher scores for the 1st born twin), this means we predict that most of the differences will be positive
- N_+ = number of positive = 7
- N_- = number of negative = 4
- n_d = number of pairs with useful info = 11

The Sign Test

$B_s = N_+ = 7$ (because of directional alternative)

$P > 0.10$, Fail to reject H_o

CONCLUSION: These data show that the aggressiveness of 1st born twins is not significantly greater than the 2nd born twins ($P > 0.10$).

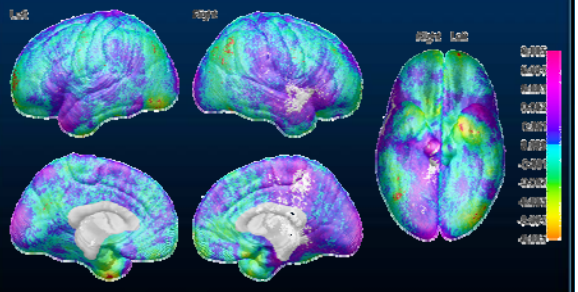
X~B(11, 0.5)

P(X>=7)=0.2744140625

http://socr.stat.ucla.edu/htmls/SOCR_Distributions.html (Binomial Distribution)

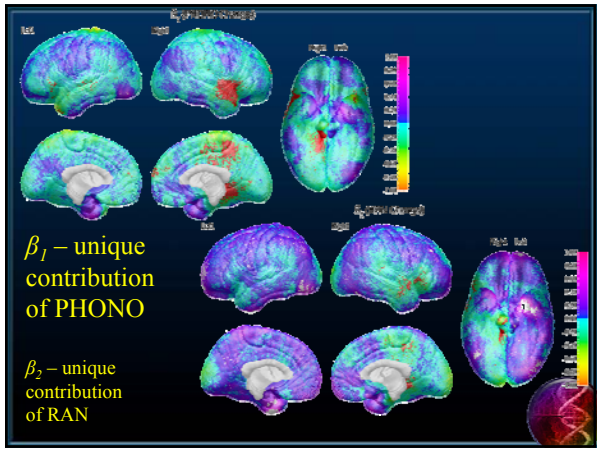
http://socr.stat.ucla.edu/Applets.dir/Normal_T_Chi2_F_Tables.htm

Phonological vs. Rapid-Naming differences using fMRI



$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

↑ Thickness Δ ↑ PHONO Δ ↑ RAN Δ ↑ interaction



β_1 – unique contribution of PHONO

β_2 – unique contribution of RAN

Approximation of the Fisher Sign Test using the normal distribution

Left ROIs	Pos	Neg	Total	Z	p
Lateral dorsofrontal	7,601	481	8,082	79	0
Lateral ventrofrontal	12,934	1366	14,300	97	0
Lateral parietal	7,659	1,701	9,361	62	0
Lateral occipital	2,905	475	3,381	42	0
Temporal	13,083	252	13,336	111	0
Medial dorsofrontal	3,484	36	3,520	58	0
Medial ventrofrontal	3,864	762	4,627	46	0
Medial parietal	3,369	199	3,568	53	0
Medial occipital	267	353	620	-3.45	<0.002

Lu, L.H., Leonard, C.M., Dinov, I.D., Thompson, P.M., Kan, E., Jolley, J., Toga, A.W., & **Sowell**, E.R. (2006, February). Differentiating between phonological processing and rapid naming using structural MRI. Paper presented at the 34th Annual Meeting of the International Neuropsychological Society, Boston, MA.

CLT

Sampling Distribution of the Sample Mean

Using the Sample Mean

Let $X_1,\ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then

1. $E(\overline{X}) = \mu_{\overline{X}} = \mu$
2. $V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n$

In addition, with $T_o = X_1 + \ldots + X_n$, $E(T_o) = n\mu$, $V(T_o) = n\sigma^2$, and $\sigma_{T_o} = \sqrt{n}\sigma$.

Normal Population Distribution

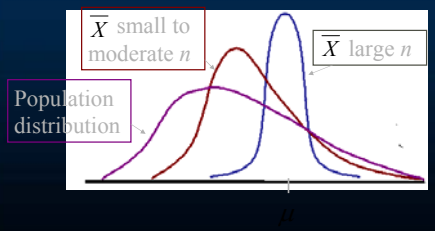
Let $X_1,\ldots, X_n$ be a random sample from a normal distribution with mean value μ and standard deviation σ Then for any n, $\overline{X}$ is normally distributed, as is T_o.

http://www.socr.ucla.edu/Applets.dir/SamplingDistributionApplet.html

The Central Limit Theorem

Let $X_1,\ldots, X_n$ be a random sample from a distribution with mean value μ and variance σ^2. Then if n sufficiently large, $\overline{X}$ has approximately a normal distribution with $\mu_{\overline{X}} = \mu$ and $\sigma^2_{\overline{X}} = \sigma^2/n$, and T_o also has approximately a normal distribution with $\mu_{T_o} = n\mu$, $\sigma_{T_o} = n\sigma^2$. The larger the value of n, the better the approximation.

The Central Limit Theorem



Central Limit Theorem - heuristic formulation

Central Limit Theorem:

When sampling from almost any distribution,

$\overline{X}$ is approximately Normally distributed in large samples.

Show Sampling Distribution Simulation Applet:
http://www.socr.ucla.edu/Applets.dir/SamplingDistributionApplet.html

Central Limit Theorem - theoretical formulation

Let $\{X_1, X_2, \ldots, X_k, \ldots\}$ be a sequence of independent observations from one specific random process. Let and $E(X) = \mu$ and $SD(X) = \sigma$ and both be finite ($0 < \sigma < \infty$; $|\mu| < \infty$). If $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, sample-avg,

Then $\overline{X}$ has a distribution which approaches $N(\mu, \sigma^2/n)$, as $n \to \infty$.

Recall we looked at the sampling distribution of $\overline{X}$

- For the sample mean calculated from a random sample, $E(\overline{X}) = \mu$ and $SD(\overline{X}) = \sigma/\sqrt{n}$, provided $\overline{X} = (X_1+X_2+ \ldots + X_n)/n$, and $X_k \sim N(\mu, \sigma)$. Then
- $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. And variability from sample to sample in the ***sample-means*** is given by the variability of the individual observations divided by the square root of the sample-size. In a way, averaging decreases variability.

Law of Large Numbers (LLN)

The **weak law of large numbers** states that if $X_1, X_2, X_3, \ldots$ is an infinite sequence of random variables, where all the random variables have the same expected value μ and variance σ^2; and are uncorrelated (i.e., the correlation between any two of them is zero), then the sample average

$$\overline{X_n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

converges in probability to μ. Somewhat less tersely: For any positive number ε, no matter how small, we have

$$\lim_{n \to \infty} P\left(\left|\overline{X_n} - \mu\right| < \varepsilon\right) = 1$$

Proof by Chebyshev's inequality!

LLN Activity:
http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_LawOfLargeNumbers