Psych 2111
Prof. Matt Jones

Lecture 10: Sampling Distributions

Same Thought Experiment
- Known population
- Sample n members
- Compute some statistic
- What is probability distribution of the statistic?
- Replication
  - Doing exactly the same experiment but with a new sample
  - Sampling variability means each replication will result in different value of statistic

Sampling Distributions
- Sampling distribution
  - The probability distribution of some statistic over repeated replication of an experiment
  - Distribution of sample means: the probability distribution for M

Reliability of the Sample Mean
- How close is M to μ?
  - Tells how much we can rely on M as estimator of μ
- Standard Error (SE or s_M)
  - Typical distance from M to μ
  - Standard deviation of $p(M)$
- Estimating μ from M
  - If we know M, we can assume μ is within about 2 SEs
  - **If** μ were further, we probably wouldn't have gotten this value of M
- SE determines reliability
  - Low SE → high reliability; high SE → low reliability
  - Depends on sample size and variability of individual scores
- Law of Large Numbers
  - The larger the sample, the closer M will be to μ
  - Formally: as n goes to infinity, SE goes to 0
  - Implication: more data means more reliability

Central Limit Theorem
- Characterizes distribution of sample mean
  - Deep mathematical result
  - Works for any population distribution
- Three properties of $p(M)$
  - Mean: The mean of $p(M)$ always equals μ
  - Standard error: The standard deviation of $p(M)$, s_M, equals $\sigma/\sqrt{n}$
- Variance equals σ^2/n
  - Shape: As n gets large, $p(M)$ approaches a Normal distribution
- All in one equation: $p(M) \approx Normal\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- Only Normal if n is large enough!
  - Rule of thumb: Normal if $n \geq 30$

Distribution of Sample Variances

Same story for s as for M

Probability distribution for s over repeated replication

Chi-square distribution (χ^2)

Probability distribution for sample variance

Positive skew; variance sensitive to outliers

Mean equals σ^2, because s^2 is unbiased

Recall: $s^2 = \dfrac{\sum (X - M)^2}{n-1}$

Distribution of $\dfrac{\sum (X - M)^2}{n}$?