The Normal Distribution

The *Standard* Normal Distribution, $z$ –scores, and Confidence Intervals

Introduction

• The Normal Distribution
  – “Prototype” used for most statistical tests
  – We can describe the “shape” of our distributions using terms such as “modality, kurtosis, and skew”
    • BUT, in “real-world” terms—what does the shape of the distribution mean?
    • How can the shape of the distribution be used to predict or describe behavior?
  – Since the “line” itself means very little, we have to expand our understanding of distributions
How Distributions Work

1. There is a critical link between the shape of a distribution for a given behavior and the probability of that behavior occurring
   - The *area* under the curve may be translated to a probability
     - area = chance of a behavior or event occurring

Self-esteem

Soc. Support

Intelligence

• Self-esteem = 50%
• Social Support = 48%
• Intelligence = 2%
How Distributions Work

2. Furthermore, we can explore the grouping, or addition of two or more areas
   • Two areas, combined, becomes one single, larger area with a similarly larger probability

- Self-esteem = 50%
- Social Support + Intelligence = 48%
Normal Distributions

• In statistics, we tend to be most interested in the normal distribution
  1. It is often safe to assume that our DVs are normally distributed in the general population
     • Oftentimes, normal distribution of the DV is an assumption, or rule, of our statistical tests
  2. It is relatively easy to calculate the probability associated with normally distributed variables
     • Possible, but much more complex, with non-normal variables (i.e. calculus)

Normal Distributions

• In statistics, we tend to be most interested in the normal distribution
  3. When large samples are drawn from populations
     1. The shape of the sample distribution is, approximately, normally distributed
     2. Large samples = more normal distribution of values
Normal Distribution

• It is mathematically possible to calculate probabilities for any point on the normal distribution from a given mean and S.D.
  – Can you see any weaknesses to this approach?

• Must calculate a new probability for EVERY change in the mean and S.D.
  – 50 vs. 51
  – 50 vs. 50.00001

Standard Normal Distribution

• Solution: create a single normal distribution with all of the associated probabilities for points on the line pre-determined
  – Thus, the standard normal distribution is a normal distribution with the mean and standard deviation arbitrarily set
    • $\bar{X} = 0$
    • $\sigma = 1$
So what?

• I know what you’re thinking…
  – We very rarely encounter distributions with means and S.D.s equal to these predetermined values
    • Further, what’s the probability of finding BOTH at the same time!
    • So, the standard normal distribution is worthless, right?
  • Maybe not… how can we use the standard normal distribution to our advantage?
Standardized Scores

- We can mathematically manipulate, or *transform*, our data
  - Convert it to a form that possesses a mean equal to 0 and a S.D. equal to 1
    1. Subtract the mean of the sample from EACH data point
    2. Divide each data point by the S.D. of the sample
  - This process converts our *raw scores* to *standardized scores* (*z*-scores)

\[ z = \frac{(x - \mu)}{\sigma} \]
### Standardized Scores

\[ z = \frac{(x - \mu)}{\sigma} \]

### Table

<table>
<thead>
<tr>
<th>Score</th>
<th>Minus mean</th>
<th>Divided by S.D</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>2.5</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>3.5</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

\[ \mu = 5 \quad \sigma = 2 \]
Standardized Scores

• We can convert ANY normally distributed variable to standardized scores
  – HOWEVER, converting to standardized scores will NOT change the shape of the distribution
    • If it was non-normal before, it will remain so after conversion to standardized scores
  
• Since we refer to standardized scores as z-scores, we sometimes refer to the standard normal distribution as the z-distribution

Standardized Scores

• Utility of standardized scores
  – Since we can convert any normally distributed variable to standardized scores...
    • We can then use the pre-generated probabilities associated with points on the curve of the standard normal distribution to estimate the probability associated with finding that particular value

• Tabulated values of the standard normal distribution (p. 759)
Calculating Probabilities for Values on the Standard Normal Distribution

• The standard normal distribution

![Diagram of the standard normal distribution with shaded areas representing different probabilities.]

Calculating Probabilities for Values on the Standard Normal Distribution

• Typically, we are interested in calculating three probabilities associated with the standard normal distribution
• When we do so, we are calculating probabilities for the total area described
  – Mean to z-score
  – The largest portion of the curve in terms of a given z-score
  – The smallest portion of the curve in terms of a given z-score
• From these 3 probabilities, several other probability calculations may be performed
Calculating Probabilities for Values on the Standard Normal Distribution

• The Mean to $z$

Calculating Probabilities for Values on the Standard Normal Distribution

• Largest portion
Calculating Probabilities for Values on the Standard Normal Distribution

• Smallest portion

Calculating Probabilities for Values on the Standard Normal Distribution

• As indicated before, the areas under the curve represented in graphs presented before map to probabilities
  – In order to calculate the probability associated with any single raw score we need do the following
    1. Convert raw score to z-score
    2. Look up z-score on a z-table (p. 759)
z-score Probabilities

• Given a z-score of .5, what is the probability of:
  – The mean to z?
    • .1915
  – The largest portion of the curve?
    • .6915
  – The smallest portion of the curve?
    • .3085

z-score Probabilities

• Notice:
  1. Mean to z + smaller portion = .50
  2. Largest portion + smaller portion = 1.00
• You can always check if your conclusions make sense
• Also, you can add and subtract probabilities to answer more complex questions
z-score Probabilities

- Given a z-scores of .5 & -1.5, what is the probability of finding a score between those two values:
  - The mean to z for .5?
    • .1915
  - The mean to z for -1.5?
    • .4332
  - Mean to z for .5 + mean to z for -1.5
    • .6247
z-score Probabilities

• Given a z-scores of .5 & 1.5, what is the probability of finding a score between those two values:
  – The mean to z for .5?
    • .1915
  – The mean to z for -1.5?
    • .4332
  – Mean to z for 1.5 - mean to z for .5
    • .2417
z-score Probabilities

- It is also possible to determine which z-score would be associated with a given probability (percentage of area under the curve)
- What z-score accounts for 10% of the area between the mean and z?
  - The z-score for .10?
    - ~ .25

Full Example

- Intelligence is typically measured with an IQ test (WAIS or WISC).
  - μ = 100
  - σ = 15
- You are tasked with determining if a given individual shows exceptional intellectual functioning.
  - IQ = 147
Convert to z-score

\[ z = \frac{（x - \mu）}{\sigma} \]

\[ \sigma = 15 \]

\[ z = \frac{（147 - 100）}{15} \]

\[ z = \frac{47}{15} \]

\[ z = 3.13 \]

Determine Probability of z

• \[ z = 3.13 \]
  - Mean to \[ z \]
    • .4993
  - Largest Portion
    • .9993
  - Smallest Portion
    • .0007

• Which score do we use to evaluate this question?
  - Smallest portion

• What represents an “exceptional” or “infrequent” score?
Identifying “Extreme” Scores

- In psychology, we tend to assume that scores less than fall outside of the most common 95% of the scores are extreme, infrequent, or exceptional
  - Area between z-score $\sim +/\sim 1.96$
  - Mean to z $1.96 +$ mean to $z -1.96 = \sim .9500$
Identifying “Extreme” Scores

• Consequently, a z-score of 3.13 is very infrequently found in the population
  – The score is exceptional

Identifying “Extreme” Score

• Thus, we assume that any z-score with a value greater than 1.96 or less than -1.96 is found infrequently
  – These are extreme scores
• Generally, this is the process we follow in calculating statistical tests
  – Determine which values are expected
  – Test if the values we possess fall outside of these boundaries
95% Confidence Interval

To make our lives easier, it is often beneficial to calculate which raw-score values fall at the boundaries of the commonly found scores

- We call this statistic the 95% confidence interval
  - Defined as the range of values for which we expect 95% of the total raw scores to fall within

\[ x = \mu \pm 1.96\sigma \]

Thus, the 95% CI is comprised of two values—an upper-bound and a lower-bound—by calculating two values for \( x \)
- Subtracting 1.96\( \sigma \) from \( \mu \) once
- Adding 1.96\( \sigma \) to \( \mu \) once
95% Confidence Interval

- For IQ:
  \[ x = \mu \pm 1.96\sigma \]
  \[ x = 100 - 1.96(15) \]
  \[ x = 100 - 29.4 \]
  \[ x = 70.6 \]
  \[ x = 100 + 29.4 \]
  \[ x = 129.4 \]

- Thus, the 95% CI for IQ is 70.6 to 129.4

Things to Remember

- 1.96 represents the z-score at which 2.5% of the total scores fall above
- -1.96 represents the z-score at which 2.5% of the total scores fall below
- 95% of the scores fall between -1.96 & 1.96

- Why might 1.645 & -1.645 be important to know?