Vocabulary

Define the following terms. Use your own words as much as possible.

1. Mean: ____________________________________________________________

2. Operational definition: ____________________________________________

3. Tail: _____________________________________________________________

4. Quartile: _________________________________________________________

5. Standardized distribution: __________________________________________

Conceptual questions

1. Write the measurement scale of each variable.
   - Number of hats a person owns: ______________
   - Type of footwear (sneakers, sandals, boots, barefoot, etc.): ______________
   - Quantity of orange juice a person can consume in 5 minutes: ______________
   - Temperature in Fahrenheit: ______________

2. Write the preferred measure of central tendency for each situation.
   - Nominal-scale variable: ______________
   - Interval-scale variable with skewed distribution: ______________
   - Ratio-scale variable with normal distribution: ______________
   - Ordinal-scale variable: ______________
   - A person’s favorite ice cream flavor: ______________

3. In the variability lecture I explained that sample range isn’t a good statistic, because it tends to get larger as sample size increases. In particular, sample range will always be less than population range (unless the sample happens to contain the minimum and maximum scores in the population). This makes sample range a(n) _____________________________ of population range.
4. Hypothesizing that being barefoot makes people less thirsty, you test how much orange juice each of your subjects drinks in a 5-minute period. Half of the subjects are tested immediately when they walk into the lab, and half are tested after sitting barefoot in the lab for 12 hours. You find the first group drinks more juice on average, but unfortunately that’s because they were tested in the morning whereas the second group was tested 12 hours later, and of course people just don’t like to drink orange juice at night.

What is the independent variable? 

What is the dependent variable? 

What is the confounding variable? 

What would you call the first (non-barefoot) group? 

Imagine you chose your groups by convenience: People wearing sandals were assigned to the barefoot group because they were more willing to take them off, whereas people wearing heavier shoes or boots were assigned to the other group. What property of good experiment design does this violate?

5. Here’s a standard normal distribution. Add something to the picture that indicates the probability that someone’s z-score will be greater than 2.

Now look at what you drew and try to estimate what that probability is. You don’t have to be right; just try to give a sensible guess.

\[ P(Z \geq 2) = \] 

Math questions

Here’s a z table, showing the probability of a z-score greater than each listed value (in a normal distribution). Use this table for Questions 1-3.

<table>
<thead>
<tr>
<th>z</th>
<th>p(Z ≥ z)</th>
<th>z</th>
<th>p(Z ≥ z)</th>
<th>z</th>
<th>p(Z ≥ z)</th>
<th>z</th>
<th>p(Z ≥ z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.500</td>
<td>.250</td>
<td>.401</td>
<td>.500</td>
<td>.309</td>
<td>.750</td>
<td>.227</td>
</tr>
<tr>
<td>.025</td>
<td>.490</td>
<td>.275</td>
<td>.392</td>
<td>.525</td>
<td>.300</td>
<td>.775</td>
<td>.219</td>
</tr>
<tr>
<td>.050</td>
<td>.480</td>
<td>.300</td>
<td>.382</td>
<td>.550</td>
<td>.291</td>
<td>.800</td>
<td>.212</td>
</tr>
<tr>
<td>.075</td>
<td>.470</td>
<td>.325</td>
<td>.373</td>
<td>.575</td>
<td>.283</td>
<td>.825</td>
<td>.205</td>
</tr>
<tr>
<td>.100</td>
<td>.460</td>
<td>.350</td>
<td>.363</td>
<td>.600</td>
<td>.274</td>
<td>.850</td>
<td>.198</td>
</tr>
<tr>
<td>.125</td>
<td>.450</td>
<td>.375</td>
<td>.354</td>
<td>.625</td>
<td>.266</td>
<td>.875</td>
<td>.191</td>
</tr>
<tr>
<td>.150</td>
<td>.440</td>
<td>.400</td>
<td>.345</td>
<td>.650</td>
<td>.258</td>
<td>.900</td>
<td>.184</td>
</tr>
<tr>
<td>.175</td>
<td>.431</td>
<td>.425</td>
<td>.335</td>
<td>.675</td>
<td>.250</td>
<td>.925</td>
<td>.177</td>
</tr>
<tr>
<td>.200</td>
<td>.421</td>
<td>.450</td>
<td>.326</td>
<td>.700</td>
<td>.262</td>
<td>.950</td>
<td>.171</td>
</tr>
<tr>
<td>.225</td>
<td>.411</td>
<td>.475</td>
<td>.317</td>
<td>.725</td>
<td>.234</td>
<td>.975</td>
<td>.165</td>
</tr>
</tbody>
</table>
1. Imagine you have a normally distributed variable with a mean of 50 and a variance of 16. What proportion of scores are greater than or equal to 53?

2. IQ is approximately normally distributed with a mean of 100 and a standard deviation of 15. What’s its interquartile range?

3. In an IQ sample of 1000 people, what would you expect the cumulative frequency to be for an IQ of 109?

4. Calculate the standard deviation of the sample {27, 36, 31, 29, 32}.

5. Think of a standard die, of the kind used in board games. The sides are numbered 1 through 6, and each number is equally likely to come up. What’s the expected value of the number you’ll get on any single roll?
**R questions**

1. List three R functions, showing an example command you would enter (don’t just write the name of the function) and a verbal description of what the command does.

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; mean(X)</td>
<td>computes the mean of a vector X</td>
</tr>
</tbody>
</table>

2. What is being computed by the following R commands? Be as precise as possible. (It may help to write what step1 and step2 are next to the first two lines.)

```
> step1 = sum(X)/length(X)
> step2 = (X - step1)^2
> answer = mean(step2)
```

3. What is being computed by this command?

```
> max(X) - min(X) + 1
```

4. What is the result of the following commands? (It may help to write out what X, Y, and Z are.)

```
> X = 101:103
> Y = 1:3
> Z = X - Y
> Z[1]
```

5. Using the z table above, what is the result of the following command? (Remember that by default `pnorm` uses the lower tail.)

```
> pnorm(.525)
```