Exam 1
Psych 3101, Fall 14

Vocabulary

1. Mean: A measure of central tendency, defined (when there are finitely many scores) as the sum of the scores divided by the number of scores
2. Operational definition: The procedure by which a construct is measured for a particular study
3. Tail: Either end of a distribution, containing the highest or lowest scores
4. Quartile: A value of a variable that is greater than 1, 2, or 3 quarters of the scores in a distribution
5. Standardized distribution: A distribution of z-scores; a distribution that has been transformed to have a mean of 0 and a standard deviation of 1

Conceptual questions

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

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

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) biased estimator 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? Barefoot vs. shod

What is the dependent variable? Quantity of orange juice consumed

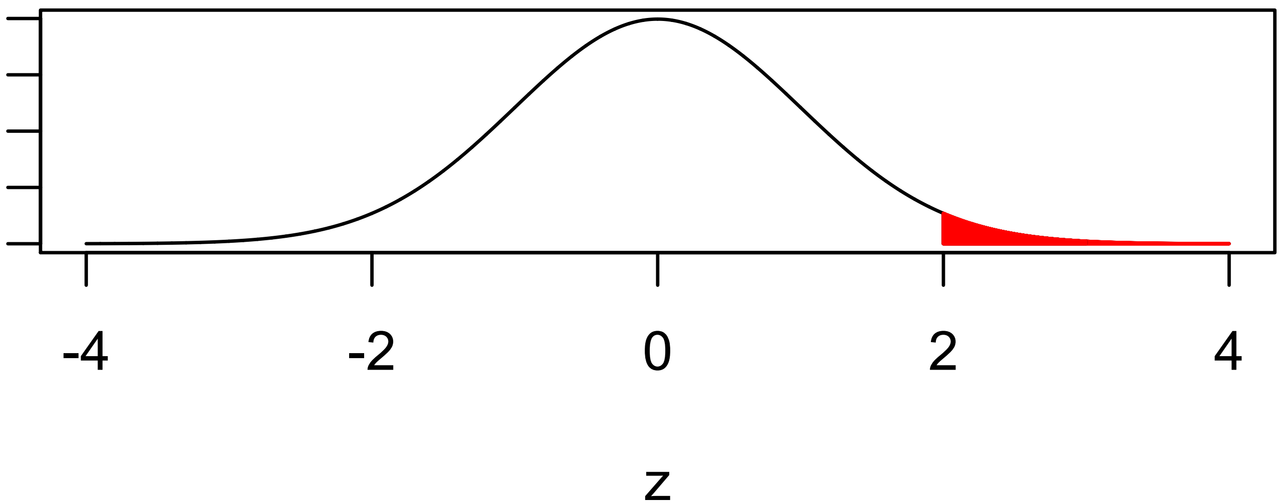
What is the confounding variable? Time of day

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

Random assignment

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.

$P(Z \geq 2)$ = .02

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.

z	p(Z≥z)	z	p(Z≥z)	z	p(Z≥z)	z	p(Z≥z)
0	.500	.250	.401	.500	.309	.750	.227
.025	.490	.275	.392	.525	.300	.775	.219
.050	.480	.300	.382	.550	.291	.800	.212
.075	.470	.325	.373	.575	.283	.825	.205
.100	.460	.350	.363	.600	.274	.850	.198
.125	.450	.375	.354	.625	.266	.875	.191
.150	.440	.400	.345	.650	.258	.900	.184
.175	.431	.425	.335	.675	.250	.925	.177
.200	.421	.450	.326	.700	.262	.950	.171
.225	.411	.475	.317	.725	.234	.975	.165

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?

$$\sigma = \sqrt{\sigma^2} = \sqrt{16} = 4$$

$$Z = \frac{53 - 50}{4} = .75$$

$$P\left[z \geq .75\right] = \underline{.227}$$

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

The table indicates the third quartile is at z = .675, because P[$z \geq .675$] = .250 = 1/4. Because the normal distribution is symmetric, the first quartile is at z = -.675.

Third quartile:

$$.675 = \frac{X - 100}{15}$$
$$X = 15 \cdot .675 + 100 = 110.13$$

First quartile:

$$-.675 = \frac{X - 100}{15}$$
$$X = 15 \cdot (-.675) + 100 = 89.87$$

Interquartile range = 110.13 – 89.87 = 20.26

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

$$Z = \frac{109 - 100}{15} = .6$$
$$P[z \leq .6] = 1 - P[z \geq .6] = 1 - .274 = .726$$
$$F(109) = 1000 \cdot .726 = \underline{726}$$

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

$$M = \frac{27 + 36 + 31 + 29 + 32}{5} = 31$$
$$s = \sqrt{\frac{\sum (X - M)^2}{n - 1}} = \sqrt{\frac{(-4)^2 + 5^2 + 0^2 + (-2)^2 + 1^2}{4}}$$
$$= \sqrt{\frac{16 + 25 + 0 + 4 + 1}{4}} = \sqrt{\frac{46}{4}} \approx \underline{3.39}$$

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?

Each outcome has probability 1/6

$$\sum_x x \cdot p(x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = \underline{3.5}$$

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.

	Command	Description
Example	`> mean(X)`	computes the mean of a vector X
1	`> log(X)`	calculates the logarithm of X
2	`> pnorm(X)`	tells the probability of a z-score less than X, in a normal distribution
3	`> prod(X)`	multiplies together the elements of a vector X

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)`	mean of X
`> step2 = (X - step1)^2`	squared deviations
`> answer = mean(step2)`	average squared deviation

population variance

3. What is being computed by this command?

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

range (for a discrete count variable, or for a continuous variable that was rounded to whole numbers)

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

`> X = 101:103`	`X = 101  102  103`
`> Y = 1:3`	`Y =   1    2    3`
`> Z = X - Y`	`Z = 100  100  100`
`> Z[1]`	First element of `Z`

`100`

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)
```

Probability of a z-score less than .525. Using the z table: 1 - .300 = .700