POLS 203
Social Statistics
Fifth Discussion Session - Statistics

Berk Filcan

November 26, 2020

1 Principles of Scientific Inquiry

What Characterizes Scientific Inquiry

Question 1

Fill in the blanks with the proper term belonging that description.

- [ ] : This is the extent to which you can be confident that a cause-and-effect relationship established in a study cannot be explained by other factors. How good your independent variable measures your dependent variable.
- [ ] : It is about how applicable a research is to the real world. How the results can be applied to other situations than your research.
- [ ] : This is a variable that influences both the dependent variable and independent variable, causing a spurious association.

Question 2

Give an example of confounding variable why below situations are spurious relationships.

- Drowning rates rises in Antalya when ice cream sales rise: [ ]
- A student works as a nurse at a local hospital. He noticed that the patients who received radiation therapy were also those most likely to die: [ ]
- A country with a high rate of healthy citizens is more likely to invent the latest technology: [ ]

What Do Statisticians Do

Question 3

Fill in the blanks with the proper term belonging that description.

- [ ] : Planning data collection and best method to study it
- [ ] : Summarizing the data
- [ ] : Making predictions using the data

Fundamental Concepts

Question 4

Fill in the blanks with the proper term belonging that description.

- [ ] : The observations gathered on the characteristics of interest
- [ ] : Total set of observations
- [ ] : A subset of the population on which one collects data
- [ ] : A measurable characteristic of a population, e.g. mean or standard deviation. It tells us something about the whole population.
- [ ] : These are numbers that summarize data from sample.

Question 5

Find the answers according to this question:
The main campus at Penn State University has a population of approximately 42,000 students. A research question is "what proportion of these students smoke regularly?" A survey was administered to a sample of 987 Penn State students. Forty-three percent (43%) of the sampled students reported that they smoked regularly. How confident can we be that 43% is close to the actual proportion of all Penn State students who smoke?

- Population: [ ]
- Parameter of interest: [ ]
- Sample: [ ]
- Statistic: [ ]

2 Sampling and Measurement

Variable Types and Measurement Scales

Question 6

Fill the table according to the measurement scale (nominal, ordinal, interval), and the type (quantitative v. categorical, discrete v. continuous) of the given variables.

	Quantitative(Numeric)		Categorical(Qualitative)	
	Discrete	Continuous	Nominal	Ordinal
Birth Year				
Gender				
T-Shirt Size				
Height				
Likert Scale				
Favorite Color				
Temperature				

Biases

Question 7

Sort below examples according to type of bias.

- Sampling Bias:
- Response Bias:
- Missing Data:

1. When the Brodie helmet was introduced during WWI, there was a dramatic rise in field hospital admissions of severe head injury victims.This led army command to consider redrawing the design, until a statistician remarked that soldiers who might previously have been killed by certain shrapnel hits to the head (and therefore never showed up in a field hospital), were now surviving the same hits, and thus made it to a field hospital.

2. In a survey utilizing a Likert scale with potential responses ranging from one to five, the respondent may only give answers as ones or fives.

3. In a study on stress and workload, employees with high workloads are less likely to participate. The resulting sample may not vary greatly in terms of workload.

4. In the early days of opinion polling, the American Literary Digest magazine collected over two million postal surveys and predicted that the Republican candidate in the U.S. presidential election, Alf Landon, would beat the incumbent president, Franklin Roosevelt, by a large margin. The result was the exact opposite. The Literary Digest survey represented a sample collected from readers of the magazine, supplemented by records of registered automobile owners and telephone users. This sample included an over-representation of individuals who were rich, who, as a group, were more likely to vote for the Republican candidate. In contrast, a poll of only 50 thousand citizens selected by George Gallup's organization successfully predicted the result, leading to the popularity of the Gallup poll.

5. The friendship paradox is the phenomenon first observed by the sociologist Scott L. Feld in 1991 that most people have fewer friends than their friends have, on average. Or, said another way, one is less likely to be friends with someone who has very few friends. In contradiction to this, most people believe that they have more friends than their friends have.

Sampling Methods

Question 8

Sort below examples according to sampling method.

- Systematic Random Sampling:
- Stratified Random Sampling:
- Cluster Sampling:

1. For their term project, a group of psychology students choose 5 students from every department of Boğaziçi University (There are nearly 30 departments in our university, so 150 participant.).

2. Another group of students choose 30 students from every level of academic standing (preparatory, freshman, sophomore, junior, senior).

3. The last group of students choose every 5th student that gets off the shuttle between South and North Campuses.

3 Descriptive Statistics

Question 9

Sort mean, median, and mode according to the shape of the distribution.

- Right-skewed distribution: [ ] > [ ] > [ ]
- Left-skewed distribution: [ ] > [ ] > [ ]

Question 10

Which of the below statements are true?

- Median is very sensitive to outliers. [ ]
- There can be multiple modes in a sample. [ ]
- Median is always same as mode. [ ]
- Standard deviation is always equal to or greater than zero. [ ]
- To calculate the variance of a sample, the sum of squares are divided by number of observations. [ ]
- μ and σ are statistics, $\bar{y}$ and s are parameters. [ ]

4 Probability

Basic Rules of Probability

Question 11

Fill in the blanks with convenient operators(+, x, -) or numbers.

- $P(\bar{A}) =$ [ ] $- P(A)$
- If A and B are distinct possible outcomes, then:
  $P(AorB) = P(A)$ [ ] $P(B)$
- If A and B are possible outcomes, then:
  $P(AandB) = P(A)$ [ ] $P(B|A)$
- If A and B are independent, then:
  $P(AandB) = P(A)$ [ ] $P(B)$

Probability Distributions, Discrete and Continuous Variables

Question 12

You wanted to try your luck and bought a ticket for New Year's Lottery. The price of a full ticket is 100 TL and 37 million tickets are expected to be sold. The great prize is 80 million TL and only 1 ticket wins this. Secondary prize is also given to only one ticket and it is 8 million TL. Another prize is 1 million TL and it is given to 8 tickets. There's another prize of 100,000 TL for 30 tickets, and a prize of 10,000 for 100 tickets. Consider all tickets are full tickets and there are no other prices such as payback prize (amorti).

1. Fill the table according to the information above.

Type of Prize	Chance of Winning	Expected Value
80 Million		
8 Million		
1 Million		
100 Thousand		
10 Thousand		

2. Calculate your expected income from buying a ticket. Does it really worth it?

Normal Distribution and Z-Scores

Question 13

1. What is the z-score of the weight of a cow that tips the scales at 375 kg, if the mean weight for cows of her type is 500 kg, with a standard deviation of 35 kg?

$$z = \frac{value - mean}{standard\ deviation} = \frac{x-\mu}{\sigma} = \frac{\square - \square}{\square} = \square$$

2. Using the same mean and standard deviation above, what is the probability of selecting a cow between 400 kg and 475 kg?

$z_1 = \frac{x-\mu}{\sigma} = \frac{\square - \square}{\square} = \square$ and $z_2 = \frac{x-\mu}{\sigma} = \frac{\square - \square}{\square} = \square$

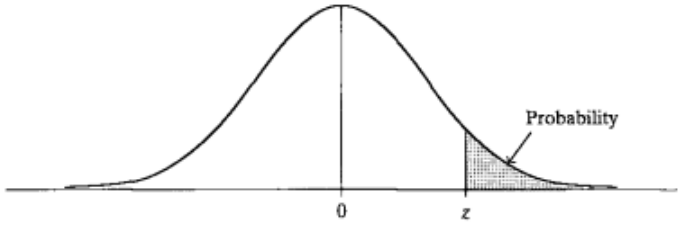
$Probability from z_1 = \square$ and $Probability from z_2 = \square$

$The\ difference\ and\ the\ answer = \square - \square = \square$

3. What is the probability of buying a cow heavier than 620 kg?

$z = \frac{x-\mu}{\sigma} = \frac{\square - \square}{\square} = \square$

$Probability from z = \square$



	Second Decimal Place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000233									
4.0	.0000317									
4.5	.00000340									
5.0	.000000287									

Source: R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).