Biostatistics 201A Fall 2021 Homework 1

Lillian Chen

10/18/2021

1. Assume Z is a standard normal random variable with mean 0 and variance 1.

(a) $P(Z \geq -0.5) =?$

Solution: $P(Z \geq -0.5) = P(Z < 0.5) \approx 0.69$

```
pnorm(0.5)
```

```
## [1] 0.6914625
```

```
pnorm(-0.5, lower.tail = F)
```

```
## [1] 0.6914625
```

(b) $P(Z <?) = 0.20$

Solution: The Z-score that corresponds to a probability of 0.20 under the standard normal distribution is $Z \approx -0.84$.

```
qnorm(0.20)
```

```
## [1] -0.8416212
```

```
qnorm(0.80, lower.tail = F)
```

```
## [1] -0.8416212
```

(c) $P(-2.0 < Z < 2.0) =?$

Solution: $P(-2.0 < Z < 2.0) = P(Z > -2.0) + P(Z < 2.0) = 2 * P(Z < 2.0) = 2 * P(Z > -2.0) \approx 1.95$

```
pnorm(-2.0, lower.tail = F) + pnorm(2.0)
```

```
## [1] 1.9545
```

```r
2*pnorm(-2.0, lower.tail = F)
```

```
## [1] 1.9545
```

```r
2*pnorm(2.0)
```

```
## [1] 1.9545
```

(d) For what value of d is it true that $P(Z < -1.5) = P(Z > d)$?

Solution: We know that the standard normal curve is unimodal and symmetric around the mode, so we can say that $P(Z < -1.5) = P(Z > d)$ is true for value of $d = 1.5$.

```r
pnorm(-1.5)
```

```
## [1] 0.0668072
```

```r
lhs <- pnorm(-1.5)
qnorm(lhs, lower.tail = F)
```

```
## [1] 1.5
```

2. Calculate the following:

(a) For a vaccine that prevents disease in 99% of the people who receive it, calculate the probability that among 1,000 people receiving the vaccine, there will be 3 or fewer people who experience the disease.
(b) What is the smallest number of independent tosses of a fair coin, n, such that the probability of obtaining either all heads or no heads is less than 0.05?

3. Assume that heights in the U.S. population are normally distributed with mean 70 inches and standard deviation 4 inches.

(a) Suppose we repeatedly take samples of size 20 from the population and calculate the sample mean height for each. What is the distribution of these sample means, and how does the standard deviation of this distribution relate to the standard deviation of heights in the population?
(b) Suppose we sample 30 people from an island and are interested to know whether the average height is significantly different from the average height in the U.S. Suppose further that the sample mean of the island individuals is 68.1 inches. Set up null and alternative hypotheses for this scenario, propose a test statistic, calculate a p-value, and comment on whether the result appears significant at the $\alpha = 0.05$ level.
(c) Following further on the example from part (b), suppose the population standard deviation is not known, but we calculate the sample variance s^2 from the sample of size 30 to be 15.0 (where, as before, the sample mean is still 68.1 inches). Again set up null and alternative hypotheses to consider whether the average height of the island population is significantly different from the average height in the U.S., propose a test statistic, calculate a p-value, and comment on whether the result appears significant at the $\alpha = 0.05$ level.
(d) Following further on the scenario in part (c), suppose we are interested to learn whether the average height of the island population is significantly different from the average height of visitors to the island airport, where a sample of size 20 visitors yields a sample mean of 70.1 inches and a sample variance of 16.2 inches. Set up null and alternative hypotheses to consider whether the average height of the island population is significantly different from the average height of island visitors, propose a test statistic, calculate a p-value, and comment on whether the result appears significant at the $\alpha = 0.05$ level.

4. Assuming a 99% confidence interval for $(\mu_1 - \mu_2)$ is 4.8 to 9.2, comment on whether each of the following conclusions should be supported or not, and explain your reasoning.

(a) Do not reject $H_0 : \mu_1 = \mu_2$ at the $\alpha = 0.05$ level if the alternative is $H_A : \mu_1 \neq \mu_2$.
(b) Reject $H_0 : \mu_1 = \mu_2$ at the $\alpha = 0.01$ level if the alternative is $H_A : \mu_1 \neq \mu_2$
(c) Reject $H_0 : \mu_1 = \mu_2$ at the $\alpha = 0.01$ level if the alternative is $H_A : \mu_1 < \mu_2$.
(d) Do not reject $H_0 : \mu_1 = \mu_2$ at the $\alpha = 0.01$ level if the alternative is $H_A : \mu_1 \neq \mu_2$.
(e) Do not reject $H_0 : \mu_1 = \mu_2 + 3$ at the $\alpha = 0.01$ level if the alternative is $H_A : \mu_1 \neq \mu_2 + 3$

5. Suppose that $\bar{X}_1 = 125.2$ and $\bar{X}_2 = 125.4$ are the mean systolic blood pressures for two samples of workers from different plants in the same industry. Suppose further that a test of $H_0 : \mu_1 = \mu_2$ using these samples is rejected at the $\alpha = 0.001$ level. Referring to available information as appropriate, critique each of the following possible conclusions, explaining your reasoning:

(a) There is a meaningful difference (clinically speaking) in population means but not a statistically significant difference.
(b) There difference in population means is both statistically and clinically significant.
(c) There is a statistically significant difference but not a clinically significant difference in population means.
(d) There is neither a statistically significant difference nor a clinically significant difference in population means.
(e) The sample sizes in the two groups must have been small.

6. In a study of "self-efficacy" (confidence in one's capability to perform a task) pertaining to exercise, subjects were randomly assigned to one of three groups. Group A received a one-time coaching session, treadmill exercise testing, and a personal trainer three times a week for 4 weeks. Group B received only the coaching session and treadmill exercise testing. Group C received an information brochure only. Self-efficacy was measured based on the responses to a series of questionnaire items. The following self-efficacy scores were observed after four weeks:

Group A: 156, 119, 100, 170, 130, 154 Group B: 132, 105, 144, 136, 132, 159 Group C: 110, 101, 124, 106, 113, 94

(a) Perform an analysis of variance, using $\alpha = 0.05$ as a significance level.
(b) Suppose Groups A and B are thought of as "active treatment" and Group C is thought of as a "control" treatment. Provide an estimate of the mean difference between active treatment and control treatment.
(c) If one was to conduct pairwise comparisons between all pairs of group means using a Bonferroni correction, what significance level would be used for each test to ensure that the "experiment-wise error rate" (i.e., the probability of at least one false finding of significance) did not exceed 0.05?

7. In a study of respiratory function among individuals depending on smoking status (to be discussed in class), consider the following statistical summaries that emerged from available data: Smoking status n Mean Std. Deviation Never 21 82.143 30.436 Former 44 84.250 29.298 Current 7 114.429 31.900

(a) Using connections between group means and the overall mean, between group means and the between-group sum of squares, and between the formula for the standard deviation and the within-group sum of squares, construct an ANOVA table based on the available information and carry out a test of significance at the $\alpha = 0.05$ level.
(b) In contrast to the study described in Problem 6, smoking status was not randomized in this study. Describe a possible objection to inferring that differences in respiratory function across smoking-status groups is due to the individual's smoking experience. Also offer a possible rejoinder to such an objection, and based on all of the knowledge you have accumulated prior to entering this course, comment on whether you would be inclined to attribute any significant differences in respiratory function to differences in smoking status.