GE2262 Quizzes

Quiz 1

1. Which of the following is not a definition of descriptive statistics?
   A. Measure the central tendency of the data
   B. Inference about the population based on the sample
   C. Identify the patterns in the data
   D. Use tables and charts to summarize the data

> B. Descriptive statistics is about summarizing the data, while inferential statistics is about infering the population parameters based on the sample statistics.

2. Consider seven distinct population observations in an ordered array below:

```
a<b<c<d<e<f<g
```

Which of the following statements are true?

I. The mean is d
II. The median is d
III. The range is g-a
IV. There is no mode

> II, III, IV.

3. In a boxplot, if the left whisker is longer than the right whisker, the underlining distribution is most likely

> Lower whiskers is the (minimum, Q1) and upper whiskers is the (Q3, maximum). If the left whisker is longer than the right whisker, the distribution has a longer left tail, which means it is left-skewed.

4. Which one of the following statements about population and sample is not correct?
   A. Measures computed from sample data are called statistics
   B. The sample statistics are used to make inference about the population
   C. A sample is chosen from a population
   D. Measures used to describe a population are called paradoxes

> D. Measures used to describe a population are called parameters.

5. Researchers are concerned that the weight of the average HK school child is increasing implying, among other things, that children's clothing should be manufactured and marketed in larger sizes. If X kg is the weight of school children sampled in a study, then X is an example of

> A continuous random variable.

6. The median for a list of 6 pulse rates 70, 64, 80, 74, 92, 74 is

> 74.

7. Given that the midterm and final exam both have a total score of 100. If the standard deviation of midterm scores and the final exam are 8 and 6, respectively. Which one of the

two exams is more uncertain and why?

> The midterm exam is more uncertain because the standard deviation is higher.

8. In general, which descriptive summary measures can be easily approximated from a boxplot?

> Median, Q1, Q3, minimum, maximum. Also, the range (maximum - minimum) and the IQR (Q3 - Q1).

9. 150 executives were asked the most common mistake candidates make during job interviews. Six different mistakes were given. Which chart is the best for presenting the information?

> A bar chart.

10. The width of each bar in a histogram corresponds to the

> Differences between the boundaries of the class.

Quiz 2

1. If P(A) = 0.50, P(B) = 0.45, and P(A and B) = 0.36, then P(A | B) =

$P(A \cap B) = P(A \mid B) \cdot P(B)$

$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.36}{0.45} = 0.8$

2. A set of events is collectively exhaustive when it includes _ the events that can occur in an experiment.

> All.

4. In a large metropolitan area, the probabilities are 0.86, 0.35 and 0.29 that a randomly chosen family owns a colour television set, a HDTV set or both kinds of sets. What is the probability that family owns at least one kind of sets?

$P(A) = 0.86, P(B) = 0.35, P(A \cap B) = 0.29$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.86 + 0.35 - 0.29 = 0.92$

5. Two names are drawn from among the seven members of a club for the offices of vice-president and treasurer. In how many ways can this be done?

$P(7, 2) = 7 \cdot 6 = 42$

6. The Mars company says that yellow made up 20% of their plain M&M candies, brown made up another 20%, and orange, blue, and green each made up 10%. The rest were red. Assuming you had an infinite supply of M&M, if you picked four M&M in a row, what is the probability that the second red appears in the fourth draw?

$P(R) = 1 - P(Y) - P(Br) - P(O) - P(B) - P(G) = 1 - 0.2 - 0.2 - 0.1 - 0.1 - 0.1 = 0.3$

The probability:

- 3 trials, with 1 success: $C(3, 1) \cdot 0.3^1 \cdot 0.7^2 = 0.441$
- another 1 trial, with 1 success: 0.3

The total probability is $0.441 \cdot 0.3 = 0.1323$

More generally, for independent trials with the same probability of success p, the probability of the k-th success in the n-th trial is $C(n-1,k-1)\cdot p^k\cdot(1-p)^{n-k}$.

Given $n=4, k=2, p=0.3$, the probability is $C(3,1)\cdot 0.3^2\cdot 0.7^2=0.1323$.

7. The weight of an item produced on a machine is normally distributed with mean 220 g and standard deviation 10 g. Among 3500 items that are independently produced, how many percent of them are expected to weigh between 215 g and 225 g?

$W\sim N(220,10^2)$

$215\le W\le 225\Rightarrow -0.5\le \frac{W-220}{10}\le 0.5$

$P(-0.5\le Z\le 0.5)=P(Z\le 0.5)-P(Z\le -0.5)=0.6915-0.3085=0.383$

8. For a standard normal distribution, a negative value of z indicates

> the value is below the mean.

9. A company has 25 i-Pads which independently require repair on a given day. The probability that any of them will require repair on a given day is 0.008. To find the probability that exactly 3 of the i-Pads will require repair on a given day, one will use that type of probability distribution?

> Binomial.

10. Properties of the normal distribution include
A. a continuous bell-shaped distribution
B. the random variable can only be integer values
C. a discrete probability distribution
D. the random variable can assume only a finite or limited set of values

> A (Remarks: B, C, D are the properties of the discrete probability distribution.)

Quiz 3

1. Suppose the ages of students in a statistics course follow a right-skewed distribution with a mean of 23 years and a standard deviation of 3 years. If we randomly sampled 100 students, which of the following statements about the sampling distribution of the sample mean age is incorrect?
   - The standard error of the sample mean is equal to 0.3 years
   - The mean of the sample mean is equal to 23 years
   - The standard deviation of the sample mean is equal to 3 years
   - The shape of the sampling distribution is approximately normal

(Answer: C)

The standard error of the sample mean (= standard deviation of the sample mean) is $\sigma_{\bar{X}}=\frac{\sigma}{\sqrt{n}}=\frac{3}{\sqrt{100}}=0.3$ years.

2. As the sample size increases, the standard error of the sample mean
   - remains unchanged
   - may increase or decrease or remain unchanged
   - decreases
   - increases

(Answer: C)

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

3. The gasoline consumption of an engine has a normal distribution with mean 5 gallons. In four independent tests, the probability that the average gasoline consumption is over 5 gallons is
   - 0.5
   - 0
   - 1
   - none of these

(Answer: 0.5)

4. The central limit theorem states that as the sample size gets large enough,
   - the distribution of the sample data values becomes the normal distribution
   - the distribution of the population data values becomes the normal distribution
   - the sampling distribution of the sample mean becomes the normal distribution
   - the sampling distribution of the population mean becomes the normal distribution

(Answer: C)

5. A sample of 20 items from a population with an unknown population standard deviation is selected in order to determine an interval estimate of the population mean. Which one of the following statements is not necessary?
   - We must use a t distribution
   - We must assume the population has a normal distribution
   - Sample standard deviation must be used to estimate the population standard deviation
   - The sample must have a normal distribution

(Answer: D)

Since sample size is small, CLT does not apply. We must assume the population has a normal distribution, so that sample mean $\bar{X}$ has a normal distribution.

The sample itself does not have to be normally distributed.

6. Suppose a 95% confidence interval for the population mean is [1000, 2100]. To make more useful inferences from the data, it is desired to reduce the width of the confidence interval. Which one of the following will result in a reduced interval width?
   - Increase the sample mean
   - Increase the sample size and decrease the level of confidence
   - Increase the population standard deviation
   - Increase the level of confidence and decrease the sample size

(Answer: B)

Confidence interval $\mu = \bar{X} \pm z \cdot \frac{\sigma}{\sqrt{n}}$

Increase the sample size n will decrease the standard error $\frac{\sigma}{\sqrt{n}}$.

Recall that $z_{\alpha/2}$ is the z-score such that $P(Z \geq z_{\alpha/2}) = \alpha/2$. As α decreases, $z_{\alpha/2}$ increases.

Decrease the level of confidence $1 - \alpha$ will increase α and thus decrease $z_{\alpha/2}$.

7. A random sample of 121 automobiles traveling on an interstate showed an average speed of 65 mph. From past information, it is known that the standard deviation of the population is 22 mph. The standard error of the mean is

$n = 121, \bar{X} = 65, \sigma = 22$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{22}{\sqrt{121}} = 2$

8. The standard error of the mean for a sample of 100 is 30. In order to reduce the standard error of the mean to 15, we would
   - increase the sample size to 400
   - decrease the sample size to 50
   - increase the sample size to 200
   - decrease the sample size to 25

(Answer: A)

9. The head librarian at the Library of Congress has asked her assistant for an interval estimate of the mean number of books checked out each day. The assistant provides the following interval estimate: from 740 to 920 books per day. If the head librarian knows that the population standard deviation is 150 books checked out per day, approximately how large a sample did her assistant use to determine the interval estimate?

Since level of confidence $1 - \alpha$ is not given, the sample size cannot be determined.

10. It is known that the population variance equals 484. With a 95% confidence, the sample size that needs to be taken if the desired margin of error is +/-5 is

Margin of error $E = z \cdot \frac{\sigma}{\sqrt{n}} = 5$

$n = \left(\frac{z \cdot \sigma}{E}\right)^2 = \left(\frac{1.96 \cdot 22}{5}\right)^2 = 74.3733 \approx 75$

Quiz 4

Coverage: Topic 6 (Hypothesis Testing), Topic 7 (Sample Proportion)

Fast Facts:

Decision \ Truth	H0 is True	H0 is False
Do not reject H0	Level of confidence ($1 - \alpha$)	Type II error (β)
Reject H0	Type I error (α)	Power of the test ($1 - \beta$)

- Two-tail test: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$
- Lower-tail test (lower-tail is rejection region): $H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$
- Upper-tail test (upper-tail is rejection region): $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$

Z-test: $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

t-test: $t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$

critical value approach: reject H_0 if $|Z| > Z_{\alpha/2}$ (two-tail test)

p-value approach: reject H_0 if $p = P(|Z| > |z|) < \alpha$ (two-tail test)

Sample proportion: $p = \frac{Y}{n}$

Sampling distribution of sample proportion: $\mu_p = \pi, \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

CI: $p \pm Z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$, where standard error $\sigma_p = \sqrt{\frac{p(1-p)}{n}}$, sampling error $E = Z_{\alpha/2} \cdot \sigma_p$

Sample size determination: $E = Z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} \Rightarrow n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 \cdot \pi(1-\pi)$

CLT conditions: $n \geq 30, n\pi \geq 5, n(1-\pi) \geq 5$

1. If the level of significance of a hypothesis test is raised from 0.01 to 0.05 given all other variables remain unchanged, the probability of a Type II error

Answer: will decrease

Increased α -> decreased β -> decreased Type II error.

2. If p-value = 0.01, which is the best conclusion for a hypothesis testing?

```
There is 1% probability that the alternative hypothesis is true
The null hypothesis might be false, but it’s unlikely
The null hypothesis is definitely false
The null hypothesis might be true, but it’s unlikely
```

Answer: The null hypothesis might be false, but it’s unlikely

3. The level of significance

```
can be any value between 0 to 1
can be any value
is the power of the test
can be any positive value
```

Answer: can be any value between 0 to 1

4. If at 34 degrees of freedom, the t-test statistic is −4.322, then what is the approximate of the p-value for a lower-tailed test?

```
P(t ≤ −4.322) < 0.005
P(t ≥ −4.322) > 0.995
P(t ≥ −4.322) < 0.050
P(t ≤ −4.322) + P (t ≥ 4.322) < 0.010
```

Answer: P(t ≤ −4.322) < 0.005

This implies that H0 is unlikely to be true.

Since lower-tail is rejection region, and the p-value is the probability of t in the rejection region.

5. We have created a 95% confidence interval for the population mean with the result [10, 15]. What decision will we make if we test H0: $\mu = 16$ versus H1: $\mu \neq 16$ at the 2.5% level of significance?

```
Fail to reject H0 in favor of H1
We cannot tell what our decision will be from the information given
Accept H0 in favor of H1
Reject H0 in favor of H1
```

Answer: Reject H0 in favor of H1

$\mu \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \mu \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}} = [10, 15]$

$\mu = 12.5, \frac{\sigma}{\sqrt{n}} = \frac{2.5}{1.96} = 1.2755$

test statistic: $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{12.5-16}{1.2755} = -2.732$

$p = P(|Z| > 2.732) = 0.0063 < 0.025$ (two-tail test)

Therefore, reject H_0 in favor of H_1.

Do not reject H_0.

6. The marketing manager for an automobile manufacturer is interested in determining the proportion of new compact-car owners who would have purchased a GPS navigation system if it had been available for an additional cost of $3000. The manager believes from previous information that the proportion is 0.30. Suppose that a survey of 200 new compact-car owners is selected and 80 indicate that they would have purchased the GPS navigation system. If you were to conduct a test to determine whether there is evidence that the proportion is different from 0.30 and decided not to reject the null hypothesis, what conclusion could you reach?

```
There is sufficient evidence that the sample proportion is 0.30
There is sufficient evidence that the population proportion is not 0.30
There is insufficient evidence that the population proportion is 0.30
There is insufficient evidence that the population proportion is not 0.30
```

$\pi = 0.30, n = 200, Y = 80, p = \frac{Y}{n} = 0.40$

$H_0 : \pi = 0.30, H_1 : \pi \neq 0.30$

test statistic: $Z = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.40-0.30}{\sqrt{\frac{0.30\cdot 0.70}{200}}} = 3.086$

7. The president of a university would like to estimate the proportion of the student population who owns a personal computer. In a sample of 500 students, 417 own a personal computer. Which one of the following statements is false?

```
It is possible that the 99% confidence interval calculated from the data will not
contain the proportion of the student population who own a personal computer
The parameter of interest is the proportion of the student population who own a
personal computer
A 90% confidence interval calculated from the same data would be narrower than a
99% confidence interval
None of these is false
```

Answer: None of these is false

8. Which one of the following would be an appropriate alternative hypothesis?

```
The population proportion is less than 0.35
The sample proportion is less than 0.35
The population proportion is at least 0.35
The sample proportion is at least 0.35
```

Answer: The population proportion is less than 0.35

9. We are interested in conducting a study in order to determine what percentage of voters of a state would vote for the incumbent governor. What is the minimum size sample needed to estimate the population proportion with a margin of error of +/-0.08 at 95% confidence?

$$E = Z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} = 0.08$$

$$n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 \cdot p(1-p) = \left(\frac{1.96}{0.08}\right)^2 \cdot 0.5 \cdot 0.5 = 150.0625 \approx 151$$

10. A sample of 100 fuses from a very large shipment is found to have 10 that are defective. The 95% confidence interval of defective fuses is

$$p = \frac{Y}{n} = 0.1, n = 100, Z_{\alpha/2} = 1.96$$

$$p \pm Z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} = 0.1 \pm 1.96 \cdot \sqrt{\frac{0.1 \cdot 0.9}{100}} = 0.1 \pm 0.0588 = [0.0412, 0.1588]$$

Additional Questions - Quiz 4

1.Which of the following statements is not true about the level of significance in a hypothesis test?

```
The level of significance provides the critical value(s) of the hypothesis test
The level of significance is the maximum risk we are willing to accept in making
a Type II error
The significance level determines the rejection region(s)
The larger the level of significance, the more likely you are to reject the null
hypothesis
```

Answer: The level of significance is the maximum risk we are willing to accept in making a Type II error

The larger the level of significance α, the smaller the risk of making a Type II error β. However, they are not algebraically related.

The level of significance α is the probability of making a Type I error.

2. For a given sample size and all other variables remain unchanged, if the level of significance is decreased, the power of the test

Answer: will decrease

The smaller the level of significance α, the larger the risk of making a Type II error β, and the smaller the power of the test $1 - \beta$.

This means it is more likely to not reject H_0 when H_0 is false.

3. Peter provided an analysis for the building of a multi-unit retail strip for Corning Construction using the following hypothesis: H0: Do not build the Strip Mall; H1: Build the Strip Mall. Based on his calculations of the area, and the probabilities for growth, he shows the project will be profitable and suggest the client to build the Strip Mall. The client took Peter’s advice, built the Strip Mall to a profit loss. What type of error would this represent?

Answer: Type I error

Decision is to reject H_0 when H_0 is true. This is a Type I error.

7. When determining the sample size for a proportion for a given level of confidence and sampling error, as the sample proportion is closer to 0.5, the required sample size is

Answer: larger

$n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 \cdot p(1-p)$

8. A confidence interval was used to estimate the proportion of statistics students who are females. A random sample of 72 statistics students generated the following 90% confidence interval: [0.438, 0.642]. Based on the interval above, is the population proportion of females equal to 0.60?

```
No, and we are 90% sure of it
No, the population proportion is 0.54
Yes, and we are 90% sure of it
Maybe, 0.60 is a believable value of the population proportion based on the
information above
```

Answer: Maybe, 0.60 is a believable value of the population proportion based on the information above

10. In a sample of 400 voters, 360 indicated they favor the incumbent governor. The 95% confidence interval of voters favoring the incumbent is

$p = \frac{Y}{n} = 0.9, n = 400, Z_{\alpha/2} = 1.96$

CI: $p \pm Z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}} = 0.9 \pm 1.96 \cdot \sqrt{\frac{0.9 \cdot 0.1}{400}} = 0.9 \pm 0.0294 = [0.8706, 0.9294]$

11. If a null hypothesis is rejected at the 5% level of significance, it

Answer: may or may not be rejected at the 1% level of significance

Larger α means more likely to reject H_0.

17. A survey claims that 9 out of 10 doctors recommend aspirin for their patients with headaches. To test this claim against the alternative that the actual proportion of doctors who recommend aspirin is not 0.90, a random sample of 100 doctors was selected. Suppose that the test statistic is – 2.20. Can we conclude that H0 should be rejected at the (i) 10%, (ii) 5%, and (iii) 1% level of significance

Answer: (i) Yes, (ii) Yes, (iii) No

$H_0 : \pi = 0.90, H_1 : \pi \neq 0.90$

p-value $P(|Z| > 2.20) = 0.0272$

25. Suppose we want to test H0: $\mu \geq 30$ versus H1: $\mu < 30$. Which one of the following samples with the size 36 is most likely to reject H0 in favor of H1?

```
Sample mean = 26, sample standard deviation = 9
Sample mean = 32, sample standard deviation = 2
Sample mean = 27, sample standard deviation = 4
Sample mean = 28, sample standard deviation = 6
```

test statistic: $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$

(A) $t = \frac{26-30}{9/\sqrt{36}} = -\frac{4}{3/2} = -2.6667$

(B) $t = \frac{32-30}{2/\sqrt{36}} = \frac{2}{1/3} = 6$

(C) $t = \frac{27-30}{4/\sqrt{36}} = -\frac{3}{2/3} = -4.5$

(D) $t = \frac{28-30}{6/\sqrt{36}} = -\frac{2}{1} = -2$

Since this is a lower-tail test, the sample with the smallest t value is most likely to reject H_0.

Answer: (C)

26. If an economist wishes to determine whether there is evidence that average family income in a community exceeds $25,000

Answer: an upper-tail test should be used

Exceeds 25,000 => > 25,000 => does not contain a equality sign.

Therefore $H_0 : \mu \leq 25,000, H_1 : \mu > 25,000$