# STAT303: Sections 510
# Fall 2003
# Exam #2
# **Form A**

Name: ________________________________________

Section and Lab Number: ________________________________________

1. **Please write any questions or explanations on this exam. I will read them before assigning your grade.**
2. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers **clearly on the exam**. Multiple marks will be counted wrong.
3. You will have 60 minutes to finish this exam.
4. If you are caught cheating or helping someone to cheat on this exam, you both will receive a grade of **zero** on the exam. You must work alone AND you must not talk to anyone about this exam until Friday.
5. This exam is worth 100 points and the equilavent weight of a regular exam.
6. Good luck!

1. A biased sample is one that:

   A. is too small.
   B. will always lead to a wrong conclusion.
   C. has certain groups from the population overrepresented or underrepresented.
   D. is always nonrepresentative.
   E. never used if valid results are desired.

2. The weight of a package of mints is believed to be normally distributed with mean 21.37 grams and standard deviation 0.4, $X \sim N(21.37, 0.4^2)$. What is the chance that a sample of 4 packages of mints has an average weight, $\overline{X}_4$, between 21 and 22 grams?

   A. 0.0314
   B. 0.9032
   C. 0.9426
   D. 0.9670
   E. 0.9938

3. Using the rules for shift and scale changes, if $X \sim N(3, 5^2)$ and $Y = 7 - 5X$, what is the distribution of $Y$?

   A. $N(-8, 11.2^2)$
   B. $N(-8, -5^2)$
   C. $N(-8, 25^2)$
   D. $N(8, 5^2)$
   E. $N(-8, 5^2)$

4. Suppose the IQ of children with Fetal Alcohol Syndrome is normally distributed with true mean $\mu = 70$ and true standard deviation $\sigma = 12$. What does $P(X < 50)$, where X is the IQ's of these children, actually mean?

   A. how likely an average child with FAS would have an IQ of 50 or less
   B. how likely any child with FAS would have an IQ of 50 or less
   C. how likely a sample of children with FAS would have an average of 50 or less
   D. how likely one child out of a sample of children with FAS would have an IQ of 50 or less
   E. how likely the true mean IQ of children with FAS is 50 or less

5. What affects the sampling distribution of the sample mean, $\overline{X}$?

   A. whether the sample is random or not
   B. the size of the sample, $n$
   C. the parent population distribution (the population being sampled)
   D. All of the above affect the sampling distribution of $\overline{X}$.
   E. Exactly two of the above affect the sampling distribution of $\overline{X}$.

6. There exists at least 100 years of weather data, but suppose we only took a sample of years and calculated a 95% confidence interval for the true yearly average total rainfall of (16.5,22) inches (obviously this isn't for Texas A&M!). Which of the following is best interpretation of this interval?

   A. There is a 95% probability that the true yearly average total rainfall is between 16.5 and 22 inches.
   B. If we sampled many years, 95% of the yearly totals would be between 16.5 and 22 inches.
   C. As long as we sampled enough years (at least 30), we will be 95% confident that the true yearly average total rainfall between 1.65 and 22 inches.
   D. If we took many different samples of years, about 95% of the confidence intervals created from these samples would contain the true yearly average total rainfall.
   E. If we took many different samples of years, about 95% of the yearly average totals would equal the true yearly average total rainfall.

7. Referring to the confidence interval in the last problem, (16.5,22), which of the following statements are plausible for this data (with 95% confidence)?

   A. The true yearly average total rainfall could be 20 inches.
   B. The true yearly average total rainfall could be 16 inches.
   C. The true yearly average total rainfall could NOT be 22 inches.
   D. All of the above are plausible.
   E. Only two of the above are plausible.

8. We want to know the percentage of former students that use Statistics after graduation. What is the"best" sampling scheme to apply?

   A. We ask the Association of Former Students for a list of graduates. We chose the first 1000 names on the list to interview.
   B. We put an advertisement in the 3 major newspapers in U.S. asking former students to contact us regarding a survey.
   C. We ask the Association of Former Students for a list of graduates. We took a simple random sample of 1000 persons fron the list to interview them.
   D. We send 3 students to different gates in Kyle Field before the A&M-UT game. They ask the people at the gates if they are former students and if they answer yes, we ask the question of interest. The students stop after 200 answers have been recorded.
   E. None of the above would give us the proper sample.

9. Suppose you have 3 confidence intervals for $\pi$ from the same data: 90% - (0.45,0.55), 95% - (0.4,0.6) and 99% - (0.2,0.8). If I wanted to know whether the true proportion was 60% or not, what could I conclude?

   A. Since 60% is only in the 99%, I would reject 60% as a plausible value for $\pi$ at the 1% level.
   B. Since 60% is in the 95%, I would reject 60% as a plausible value for $\pi$ at the 5 and 10% levels.
   C. Since 60% is in the 95%, I would reject 60% as a plausible value for $\pi$ at the 10% level only.
   D. 60% is plausible since we can't **guarantee** the true proportion will fall in any interval.
   E. You can't make conclusions with confidence intervals, only hypothesis tests.

10. The sample proportion of red M&M's, in bags with a total of 50, is $p_{red} \sim N(0.2, (0.057^2)$. How likely are you to get only 2 reds (so $p_{red} = 2/50 = 0.04$)? What is $P(p_{red} \leq 0.04)$?

    A. $-2.81$
    B. 0.0025
    C. 0.025
    D. 0.0218
    E. $-2.017$

11. Assuming that my samples are always random, what could I do to reduce the standard deviation of my sample means, $\sigma_{\overline{X}}$ by a factor of 4, *i.e.*, make it one fourth as large?

    A. use 2 times as much data
    B. use one fourth as much data
    C. use 4 times as much data
    D. use 8 times as much data
    E. use 16 times as much data

12. Ok, I just made these up, but suppose that the distribution of quiz grades is the following:

| X | 0 | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|---|
| p(X) | 0.10 | 0.20 | 0.25 | 0.20 | 0.15 | 0.10 |

    What is the true mean score of the quiz grades?

    A. We can only estimate the true mean.
    B. 25
    C. 24
    D. 24.5
    E. 20

13. What are the $z$ critical values, the $z_{\alpha/2}$, for a 91% confidence interval?

    A. $\pm 0.48$
    B. $\pm 1.695$
    C. $\pm 0.82$
    D. $\pm 1.34$
    E. $\pm 0.67$

14. Suppose we have a biased coin so the true probability of getting a head is actually 90%, $\pi = 0.90$. How many times must we toss it for the sampling proportion, $p$, to be approximately normally distributed?

    A. A sample size, $n \geq 30$ is always necessary.
    B. Categorical data, like this, cannot ever be approximated by a normal.
    C. The rule for categorical data says at least 10 tosses are necessary.
    D. The rule for categorical data requires we use at least 100 tosses.
    E. The rule for categorical data says 12 would be sufficient.

15. Suppose we have a population of people's ages with mean $\mu = 35$ years and standard deviation $\sigma = 2.5$ years. The proportion of people in that population that are 25 years or older is 0.3. If we take a simple random sample of size 40 from the population, what is the distribution of the sample mean age?

    A. There is not enough information to decide.
    B. Binomial with $n = 40$ and $\pi = 0.30$ since the data is highly skewed
    C. $N(0.3, 0.072^2)$
    D. possibly approximately $N(35, 0.40^2)$ since the data is highly skewed
    E. $N(35, 2.5^2)$

16. Suppose the measurements on your bathroom scale are normally distributed with true mean, $\mu$ =your true weight and true standard deviation, $\sigma = 0.25$lbs. If we construct 95% confidence intervals, which of the following COULD be true about the intervals?

    A. All of the confidence intervals contain all of the sample means.
    B. All of the confidence intervals contain your true weight.
    C. 95% of the confidence intervals contain your true weight.
    D. All of the above could be true (but obviously not at the same time).
    E. Only two of the above could ever be true.

17. The level of nitrogen oxides (NOX) in the exhaust of a particular car model varies. A company has 200 cars of this model in its fleet. We're interested in obtaining a confidence interval for the mean NOX emission level. The data don't appear to come from a normal distribution but they aren't too far from normality and no outliers are present. Assuming a sample size of 200 is large enough, why can we use the confidence interval ordinarily reserved for normal data?

    A. Statistics is a fairly flexible science. You can get away with a lot more than you think.
    B. The Central Limit Theorem says it's justified.
    C. It's justified since we're using a typical level of confidence, 95%.
    D. The confidence interval for the normal distribution is used when the data is numerical, especially when the variable is continuous such as measuring the diameter of a piston.
    E. none of the above

18. Suppose $X \sim N(50, 2.4^2)$ and $Y \sim N(30, 1.5^2)$. If we take samples of size 25 from each, what is the distribution of the difference of the means, $\overline{X}_{25} - \overline{Y}_{25}$?

    A. The shape is normal and $\mu_{\overline{X}_{25}-\overline{Y}_{25}} = 20$, but we can't calculate $\sigma_{\overline{X}_{25}-\overline{Y}_{25}}$ without knowing that X and Y are independent.
    B. $\mu_{\overline{X}_{25}-\overline{Y}_{25}} = 20$ and $\sigma_{\overline{X}_{25}-\overline{Y}_{25}} = 0.57$, but we can't say the shape is normal.
    C. $\mu_{\overline{X}_{25}-\overline{Y}_{25}} = 20$ and $\sigma_{\overline{X}_{25}-\overline{Y}_{25}} = 0.9$, but we can't say the shape is normal.
    D. $N(20, 0.9^2)$
    E. $N(20, 0.57^2)$

19. What do we mean by the term *confidence* in reference to confidence intervals?

    A. We are confident that our data is random.
    B. We are confident that our method produces intervals that contain the parameter $(1 - \alpha)100\%$ of the time.
    C. We are confident that our method produces intervals that contain the parameter $\alpha * 100\%$ of the time.
    D. We are confident that our interval contains the parameter.
    E. We are confident that our method produces intervals that contain the statistic $(1 - \alpha)100\%$ of the time.

20. What is the name of the following theorem "In repeated independent samples from any population, the sample mean of the observed values will get as close to the population mean as you choose".

    A. The Central Limit Theorem
    B. The Law of Large Numbers
    C. Bayes theorem
    D. Sampling Distribution
    E. Simpson's Paradox

1C,2D,3C,4B,5D,6D,7A,8C,9C,10B,11E
12C,13B,14D,15D,16D,17B,18A,19B,20B