

Fill in the following.

1. How do you describe/compare quantitative data graphs (histogram, stem plot, dot plot)?

**SOCS – Shape, Outliers, Center, Spread.**

You can also mention anything unusual, gaps, clusters, etc.

2. What do you use to describe the center and spread of symmetric data with no outliers?

**Mean and Standard Deviation**

3. What you use to describe the center and spread of skewed data and/or data with outliers?

**Median and IQR**

4. How do you interpret standard deviation?

**The average (typical) distance the data is from the mean.**

(Don't forget that variance is  $s^2$ )

5. How do you interpret z-score?

**The number of standard deviations the data is above (pos.) or below (neg.) the mean**

6. How do you describe a scatterplot?

**DUFS – Direction, Unusual Features, Form, Strength**

7. How do you interpret the correlation coefficient,  $r$ ?

**The strength and direction of a LINEAR association between  $x$  and  $y$ .**

8. How do you interpret the coefficient of determination,  $r^2$ ?

**The percent (fraction) of variation in the values of  $y$  that is accounted for by the least-squares regression line of  $y$  on  $x$ .**

9. How do you interpret the slope of the regression line?

**For an increase of 1 in  $x$ ,  $y$  will increase/decrease \_\_\_\_\_.**

10. How do you interpret the  $y$ -intercept of the regression line?

**If  $x$  is 0, then  $y$  is \_\_\_\_\_.**

Remember that the  $y$ -intercept sometimes doesn't make sense, because  $x = 0$  doesn't make sense.

11. How do you interpret the  $s$ , the standard deviation of the residuals?

**The average (typical) size of the errors made when using the LSR line to make predictions**

12. What does the residual tell you?

**Residual is actual  $y$  – predicted  $y$ . It tells you whether your prediction was too high/low and how far your prediction was from the actual data.**

13. What does  $\hat{y}$  stand for?

**Predicted  $y$ .**

14. What is a high leverage point?

**Any point in a scatterplot that is extreme in the  $x$ -direction**

15. What is an outlier point?

**Any point in a scatterplot that is far from the trend of the data.**

16. What is an influential point?

**A point, that if removed, will result in a “significant” change in the calculations associated with LSR. The equation of LSR line,  $r$ ,  $r^2$ , and slope.**

17. What is the difference between a cluster sample and stratified sample?

**In cluster the clusters are usually heterogeneous and you randomly select some clusters and use all of the clusters selected. In stratified the strata are usually homogeneous (for some variable) and you randomly select from every strata.**

18. How do you know if a study is an experiment?

**A treatment is done on the experimental units in order to measure a response.**

19. What is the difference between stratified random sampling and block design?

**Stratified random sampling is a sampling method. Block design is an experimental design.**

20. How do determine if two events are independent?

**$P(A \text{ and } B) = P(A)P(B)$  or  $P(A) = P(A|B)$  or  $P(B) = P(B|A)$**

21. What parameters do you use to describe a normal distribution?

**Mean and Standard Deviation.  $N(\mu, \sigma)$**

22. What parameters do you use to describe a binomial distribution?

**Number of trials and probability of success in one trial.  $B(n, p)$**

23. What parameters do you use to describe a geometric distribution?

**Probability of success in one trial.  $G(p)$**

24. When you add/subtract random variables how do you calculate the new mean?

**Add/Subtract the means.  $\mu_1 \pm \mu_2$**

25. When you add/subtract random variables how do you calculate the new standard deviation?

**Add variances and take the square root.  $\sqrt{\sigma_1^2 + \sigma_2^2}$**

26. How does a linear transformation affect the mean and standard deviation of a random variable?

**You multiply both the mean and standard deviation. You add/subtract only the mean.**

**$\mu_{AX+B} = A\mu + B$ ,  $\sigma_{AX+B} = A\sigma$**

27. How do you describe a sampling distribution?

**Shape (normal and why), center (mean) and spread (standard deviation formula, 10% condition).**

28. What sentence do you use to interpret a confidence interval?

**I am \_\_\_% confident that the interval \_\_\_ to \_\_\_ captures the true (mean, proportion, difference in means, difference in proportions) context.**

29. What sentence do you use to interpret a confidence level?

**If I took many samples of size  $n$  from the same population, then \_\_\_% of them with capture the parameter. (always include context)**

30. How do you interpret P-value?

**The probability, assuming  $H_0$  is true, that the statistic would take the value or more extreme (in direction of  $H_a$ ) than the one calculated, by chance.**

31. In a hypothesis test, if  $P\text{-value} < \alpha$  what sentence do you use in your conclusion?

**We reject the null. We have convincing evidence to say  $H_a$  (context) is true.**

32. In a hypothesis test, if  $P\text{-value} > \alpha$  what sentence do you use in your conclusion?

**We fail to reject the null. We don't have convincing evidence to say  $H_a$  (context) is true.**

33. What is a type I error?

**You reject the  $H_0$  and you shouldn't have.**

34. What is a type II error?

**You fail to reject  $H_0$  and you should have.**

35. What is the power of the test?

**Rejecting  $H_0$  correctly. This is not an error, but the probability of making the correct decision.**

36. What increases power?

**Increase sample size, increase  $\alpha$ , increase the difference between the null hypothesis and the actual value of the parameter.**

37. What is the probability of a Type I error?

**Alpha,  $\alpha$**

38. How is the probability of a type II error and power related?

**The probability of type II error is  $\beta$  and the power of the test is  $1 - \beta$ .**

39. How you analyze a mosaic plot?

**You compare the areas of the bars. The plot shows the distribution of variables in groups and also shows the size of groups you are comparing by making the bars different widths.**

40. What is a systematic random sample?

**A sampling type where a random starting point is selected and then you select every  $n$ th one.**

41. What conditions need to be satisfied in order to proceed with an inference procedure?

**Random, 10% (if sampling without replacement), and Large count**

42. What is a parameter?

**A number that describes a population**

43. What is a statistic?

**A number that describes a sample**