Quick Formula Reference and Hints

Outliers

IQR Method: IQR = $Q_3 - Q_1$

Lower fence: X< $Q_1 - 1.5\times(Q_3 - Q_1)$

Upper Fence: X> $Q_1 - 1.5\times(Q_3 - Q_1)$

Normal Distribution

Normal: $\mu \pm 2(\sigma)$ or $\mu \pm 3(\sigma)$

Classical Probability

Conditional Probability: P(A|B) = $\frac{P(A\cap B)}{P(B)}$ or $\frac{\text{Intersection}}{\text{given}}$ ∩ = intersection P(A|B) reads the probability of A given B

Hint: if you see the word **Given** in a problem, you may need to **make a 2-way table**

Mutually Exclusive: The probability of the intersection is zero because the two events cannot occur at the same time. A Venn diagram will have no overlap.

Independent: If independent the following is true:

P(A∩B) = P(A) × P(B); the P(A|B) = P(A); $\frac{P(A\cap B)}{P(B)}$ **= P(B)**

Addition Rule: P(A∪B) = P(A) + P(B) - P(A∩B); ∪ means **both** and ∪ is also used to mean

If Independent

P(A∪B) = P(A) + P(B) - P(A) × P(B)

If Mutually Exclusive

P(A∪B) = P(A) + P(B) - 0

Discrete Random Variable

X	2	4	6	8
Frequency	0.15	0.30	0.35	?

Remember: The probabilities must sum to 1
So 1-[.15+.30+.35] = **?** = **2**

Enter Data into STAT Edit and then go to STATS Calc 1-Var Stats

Mean: E(x) = $\mu_x = \Sigma x_i p_i$ (2)(.15)+(4)(.30)+(6)(.35)+(8)(.20)

Standard deviation: $\sigma = \sqrt{\Sigma(x_i - \mu)^2 p_i}$

Geometric

Formula: $(1-p)^{n-1}p$ n is the number of trials until the 1st success, p is the probability of success

Mean or E(X) = $\mu = \frac{1}{p}$ **Variance or Var(X)** $\sigma^2 = \frac{q}{p^2}$ **Standard deviation** $\sigma = \frac{\sqrt{q}}{p}$

Problem has a fixed probability & trials are independent & is looking for the FIRST success

Geometric PDF: What is the probability that the **1st** success will occur on a given trial

Geometric CDF: Use when the problem is a geometric and is asking for the first success to occur within a set number of attempts. **Key words:** among, before, no more than, etc.

1-Geometric CDF: Use when the problem is a geometric and is asking for the first success to occur after a specific number of attempts. **Key words:** among, before, no more than, etc.

Binomial

Formula: $\binom{n}{x}p^x(1-p)^{n-x}$ **Hint: look for the phrase "out of"**

n is the number of trials. **p** is the probability of success. **X** is the number of successes

Mean or E(X)=np **Variance or Var(X)** σ^2 **= np(1-p)** **Standard deviation** $\sigma = \sqrt{np(1-p)}$

Binomial PDF: Use when the problem has a fixed probability and a set number of trials and is asking for the probability of a specific number of successes. **This must be a specific number of successes and not a range of values.**

Binomial CDF: Use when the problem has a fixed probability and a set number of trials and is asking for a range of values. **Key words: less than, fewer than, no more than etc.**

1-Binomial CDF: Use when the problem has a fixed probability and a set number of trials and is asking for a range of values. **Key words: at least, more than, greater than etc.**

Normal

Formula: $z = \frac{x-\mu}{\sigma}$ **You must have** σ **the standard deviation of the population.** The problem will tell you the distribution is Normal.

Inverse Norm: Use when the distribution is normal and you are given a percent, a percentile rank or a probability. **Always draw a sketch and shade as follows:**

- **Percentile Ranks:** (shade from the left to the right)
- **Key words like more or greater, etc.** (shade from the right to the left)
- **Key words like less or smaller, etc.** (shade from the left to the right)

Normal CDF or Z-test: Use when the question is asking to calculate the probability of 1 of something. The problem will often ask for the probability of **"a"…**

Always draw a sketch and shade as follows:

- **Key words like more or greater, etc.** (shade from the right to the left)
- **Key words like less or smaller, etc.** (shade from the left to the right)
- **Key words like between** indicate that you have 2 boundaries and will be subtracting the lower boundary from the upper boundary.

Distribution of the Sample Means or the Sampling Distribution

Formula: $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ **You must have** σ **the standard deviation of the population.** The problem will tell you the distribution is Normal **or that n is greater than 30.**

Normal CDF or Z-test: The problem will ask you to find the probability for the **sample mean**, **sample average** or $\bar{x}$

T-distribution

Formula: $T = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ **You are using s the standard deviation of the sample. You do not have σ** the standard deviation of the population. The problem will tell you the distribution is Normal **or that n is greater than 30.**

T-CDF or t-test: The problem will ask you to find the probability for the **sample mean**, **sample average** or $\bar{x}$

Which Test is it?

Note: Formulas & Conditions are on the Inferences Cheat Sheet

Proportions: **All tests for proportions are Z-tests.** If the problem asks for a proportion or a percent, that is an obvious give away that the test is a test for proportions. On the other hand, if summary data is given and there is **no standard deviation given, then the test is most likely a test for proportions.**

Tests for Means: If the problem asks about a mean or $\bar{x}$ and a standard deviation is given then we know it is some sort of test for means.

Z-test for means: σ the **standard deviation of the population** must be given. It is very unlikely to have this type of problem on the test.

T-test for means: **s the standard deviation of the sample is given and σ** the standard deviation of the population is not provided

Paired T-test vs. 2-Sample T-test

Note: Formulas & Conditions are on the Inferences Cheat Sheet

A **2-sample T-test** has 2 independent random samples; sample sizes may or may not be equal. A **paired t-test** must have equal sample sizes and something must have been **tested twice** (before – after) or their must be a natural pairing such as siblings or husband and wife.

Confidence Intervals

Note: Formulas & Conditions are on the Inferences Cheat Sheet

As the confidence level **increases**, the interval gets **wider**.
As the sample size **increases**, the interval gets **narrower**.
As the standard error **increases,** the interval gets **wider**.

- For Proportions: as $\hat{p}$ gets closer to .5 the interval gets **wider**.
- For Means: as the standard deviation **increases**, the interval gets **wider.**

Sample Size Formulas

Proportions

$n = \frac{(z^*)^2\hat{p}\hat{q}}{ME^2}$ **or Margin of Error** $= z^*\sqrt{\frac{(\hat{p})(\hat{q})}{n}}$

Means

$n = \frac{(z^*)^2(\sigma)^2}{ME^2}$ **or Margin of Error** $= Z^*\frac{\sigma}{\sqrt{n}}$

Power

Power=1 - β β = the probability of a Type II Error

As Sample Size increases, power increases because $\frac{s}{\sqrt{n}}$ or $\sqrt{\frac{(\hat{p})(\hat{q})}{n}}$ gets smaller

As alpha (the level of significance increases), power decreases

Power Increases, the further $\bar{x}$ (the sample mean) is from μ (the population mean) as long as it is in the same direction indicated by the inequality of the alternative hypothesis

Misc.

Random Assignment: This is for experiments and we have this condition to help equalize the effects of variables that we were not able to control for in the design of the experiment.

Random Sample: This is for observational studies and we have this condition so that we can select a sample that is representative of the population that we want to generalize our findings to.

n is less than 10%: We have this condition because we are sampling without replacement and want the amount of dependence between observations to be negligible.

Make certain you are using population notation for your null and alternative hypothesis.

Note: Hypothesis Tests & Formulas are on the Recipe Cards

Make certain you are using sample notation for your formulas

Note: Hypothesis Tests & Formulas are on the Recipe Cards