Units 1–5: Descriptive & Probability

Master Formula & Reference Sheet

STUDENT NAME: _____ DATE: _____ GOAL: Score 5 · sidtutor.com

1. Descriptive Statistics & Outliers

Sample Mean: $\bar{x} = (\Sigma x_i) / n$

Sample Standard Deviation: $s = \sqrt{[\Sigma(x_i - \bar{x})^2 / (n-1)]}$

Interquartile Range: $IQR = Q_3 - Q_1$

1.5 × IQR Fences: Lower = $Q_1 - 1.5(IQR)$ | Upper = $Q_3 + 1.5(IQR)$

Transformations (y = a + bx): Center = a + b(Center) | Spread = |b| × (Spread)

2. Normal Model & Bivariate Regression

Standardized z-Score: $z = (x - \mu) / \sigma \Rightarrow$ Raw Value: $x = \mu + z\sigma$

LSRL Equation: $\hat{y} = a + bx$ (passes through $(\bar{x}, \bar{y})$)

Regression Slope: $b = r \times (s_y / s_x)$ | **Intercept:** $a = \bar{y} - b\bar{x}$

Residual: e = observed y – predicted $\hat{y}$ ($e = y - \hat{y}$)

Coefficient of Determination: r^2 (% of y variation explained by x)

3. Probability Rules & Combinatorics

Complement Rule: $P(A^C) = 1 - P(A)$

General Addition Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional Probability: $P(A \mid B) = P(A \cap B) / P(B)$

General Multiplication: $P(A \cap B) = P(A) \times P(B \mid A)$

Independent Events: $P(A \mid B) = P(A) \Leftrightarrow P(A \cap B) = P(A) \times P(B)$

4. Random Variables & Combinations

Expected Value (Mean): $\mu_X = E(X) = \Sigma [x_i \times P(x_i)]$

Variance: $\sigma_X^2 = Var(X) = \Sigma [(x_i - \mu_X)^2 \times P(x_i)]$

Linear Transformation: $\mu_{a+bX} = a + b\mu_X$ | $\sigma_{a+bX} = |b|\sigma_X$

Sum of Independent RVs: $\mu_{X+Y} = \mu_X + \mu_Y$ | $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$

Difference of Independent RVs: $\mu_{X-Y} = \mu_X - \mu_Y$ | $\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$ (Add!)

5. Binomial & Geometric Distributions

Binomial Distribution B(n, p) (BINS: Binary, Indep, Number fixed, Same p):

- $P(X = k) = (n \text{ choose } k)\, p^k (1-p)^{n-k}$
- Mean: $\mu_X = np$ | SD: $\sigma_X = \sqrt{[np(1-p)]}$

Geometric Distribution G(p) (BITS: Trials until 1st success):

- $P(Y = k) = (1-p)^{k-1} p$
- Mean: $\mu_Y = 1/p$ | SD: $\sigma_Y = \sqrt{(1-p)/p}$

Master Formula & Reference Sheet

6. Sampling Distributions & Central Limit Theorem (CLT)

Sample Proportion ($\hat{p}$): $\mu_{\hat{p}} = p \mid \sigma_{\hat{p}} = \sqrt{[p(1-p)/n]}$

Normality Condition: Large Counts: $np \geq 10$ and $n(1-p) \geq 10$.

Independence Condition: 10% condition: $n \leq 0.10\,N$.

Sample Mean ($\bar{x}$): $\mu_{\bar{x}} = \mu \mid \sigma_{\bar{x}} = \sigma / \sqrt{n}$

Central Limit Theorem (CLT): When sample size $n \geq 30$, the sampling distribution of $\bar{x}$ is approximately normal regardless of population shape.

7. The 8 Core AP Statistics Inference Procedures Matrix

INFERENCE PROCEDURE	PARAMETER	TEST STATISTIC FORMULA	STANDARD ERROR FORMULA	DF
1-Sample z-Test for Proportion	p (true proportion)	$z = (\hat{p} - p_0) / \sqrt{[p_0(1-p_0)/n]}$	$SE = \sqrt{[\hat{p}(1-\hat{p})/n]}$ (for CI)	N/A
2-Sample z-Test for Proportions	$p_1 - p_2$	$z = (\hat{p}_1 - \hat{p}_2) / \sqrt{[\hat{p}_c(1-\hat{p}_c)(1/n_1 + 1/n_2)]}$	$SE = \sqrt{[\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2]}$	N/A
1-Sample t-Test for Mean	μ (true mean)	$t = (\bar{x} - \mu_0) / (s / \sqrt{n})$	$SE = s / \sqrt{n}$	$n - 1$
2-Sample t-Test for Means	$\mu_1 - \mu_2$	$t = [(\bar{x}_1 - \bar{x}_2) - 0] / \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$	$SE = \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$	Tech / $\min(n_1-1, n_2-1)$
Paired t-Test (Differences)	μ_d (mean difference)	$t = (\bar{x}_d - 0) / (s_d / \sqrt{n})$	$SE = s_d / \sqrt{n}$	$n_{pairs} - 1$
Chi-Square Goodness of Fit	Distribution of 1 categorical var	$\chi^2 = \Sigma\,[(Observed - Expected)^2 / Expected]$	Expected = $n \times p_i$ (all Exp ≥ 5)	$k - 1$ (categories)
Chi-Square Indep / Homogeneity	Association of 2 categorical vars	$\chi^2 = \Sigma\,[(Observed - Expected)^2 / Expected]$	Expected = (Row Tot × Col Tot) / Grand Tot	$(r-1)(c-1)$
Linear Regression Slope t-Test	β (true pop slope)	$t = (b - 0) / SE_b$	SE_b (from computer regression output)	$n - 2$

8. Type I & II Errors and Power

- **Type I Error (α):** Reject H_0 when H_0 is actually true ("false alarm"). P(Type I) = α.
- **Type II Error (β):** Fail to reject H_0 when H_0 is false ("missed discovery").
- **Power = 1 – β:** Probability of correctly rejecting a false H_0.
- **To Increase Power:** Increase sample size n, increase α, or increase true effect size.

9. Confidence Interval General Form

```
Point Estimate ± (Critical Value) × (Standard Error)
```

Confidence Level Meaning: *"In repeated random sampling, approximately [C]% of constructed confidence intervals will contain the true population parameter [parameter in context]."*