MATH-338 Midterm 2 Cheat Sheet

THEORY

Day 14: probability density function is represented an integral with function $f(x)$. Our probability lies within the curve and is always 1. Density curve → bell curve. Z-Score allows us to have a universal standard for density curves with different scales. They are directly proportional to the standard deviation and the delta from the mean of the graph.

Day 15: unimodal: one hump, bimodal: two humps. Mean is resistant whereas the mean is subject to change. Density curves decay to histograms (integral → to Reimann Sum). Whisker plots are an effective method to determine if a data set contains outliers (data points not belonging to the sample set). Left skew: long left tail. Sloping →. Right skew: long right tail. Sloping ←

Day 16: error: since there is some error while taking sample data, we do allow for some buffer. We also do not measure exact but to a tolerance which is influenced by the buffer above. Central Limit Theorem: when population size is "large enough" $\bar{x}$ is an approximation. Higher skew and outliers suggest a larger n value.

Day 18: As $n \uparrow$, $SEM \downarrow$.

Day 19: H_1: $\mu < \mu_0 \leftarrow$ left side. H_1: $\mu \neq \mu_0 \leftarrow \mathbf{n}\ \sigma$ on both side but no middle. H_1: $\mu > \mu_0 \leftarrow$ lower.tail = TRUE. Population distribution normality $\implies$ sample population distribution normality. Matched pairs design:

Day 20: Two Independent Samples t-Test: two unrelated treatments into one numerical response variable measured in two independent groups. Two different μ_1 and μ_2. NHST approach; identify μ_i

---

FORMULAS

- $\sqcap = width \times \frac{1}{width}$ (finite curve)
- $Z = \frac{x-\mu}{\sigma}$ (z-score)
- $X \sim N(\mu\,, \sigma)$
- $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$
- $SEM = \frac{s}{\sqrt{n}}$
- $t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$
- $t = \frac{\bar{x}}{\sigma_{\bar{x}}}$
- $t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}} \sim t(K)$ [NHST]
- $df = n - 1$

- $\mathrm{IQR} = Q_3 - Q_1$
- $K = 1.5$
- Lower fence: $Q_1 - K \times IQR$
- Upper fence: $Q_3 + K \times IQR$
- $t = \frac{\Delta\bar{x} - \Delta\mu}{\frac{\Delta s}{\sqrt{n}}}$

FRAMEWORK FLOW CHART

NPHT

- Parameter is μ [population mean]. $\mu_0 = \mu_1$
- $\bar{X}$ is sample mean. Under CLT, normal distribution at μ_0 for H_0 and μ_1 for H_1.
- We accept H_0 if <u>not</u> in CR.

N-P Power Analysis

- Define parameter and its value under H_0 and H_1
- Define a test statistic and its sampling distribution under both hypotheses.
- Use α to compute critical region
- Compute power and compare to 80

One-Sample T-Statistic [NP]

- If $\mathrm{t}_{\mathrm{observed}}$ in CR, then accept H_1: $\mu = \mu_1$. Else accept H_0: $\mu = \mu_0$

Two-Tailed Test

- Take the upper and lower limit of the curve and the significance level (α) is the cut off point of being *statistically significant*. Treat as critical region. If in CR, then accept H_1. Else accept H_0.

NHST

- Define a parameter and it's value under H_0.
- Define an interval representing an inequality
- Define a test statistic and its sampling distrubution under H_0
- Compute p-value. P-Value $\leq$ sig level $\implies$ reject H_0 & accept H_1. P-Value $>$ sig level $\implies$ fail to reject H_0. Can only be $>, < \neq$.

Two-Sided Test

- **Neyman-Pearson**
- Critical region is $\frac{1}{2}$ left tail and $\frac{1}{2}$ right tail of sampling distribution under H_0. Power will $\downarrow$.
- **NHST**
- Find the "one-sided" p-value and double it.

Matched Pairs t-Test

- Paired subjects receives their respective treatment or an individual gets two treatments. Also a subset of block design.
- H_0: $\mu_\mathrm{d} = 0$ (no difference) and H_a: $\mu_\mathrm{d} \neq 0$ (difference).
- Requirements: large population, normal distribution, σ is unknown.