I posted rules, formulas, tables, and last semester’s exam 2 (which excluded §7.3, 7.4, and 7.5) in the line of the syllabus.

7 Inferences for

7.1 -Sample Inferences on the Difference Between Two Population

7.2 Inferences on the Difference Between Two

7.3 -Sample Inferences on the Difference Between Two

7.4 Inferences Using Data

7.5 The $F$ Test for Equality of

7.1 Large-Sample Inferences on the Difference Between Two Population Means

We compare two population means, $\mu_X$ and $\mu_Y$, by studying their difference, $\mu_X - \mu_Y$. Notation:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Population 1</th>
<th>Population 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$\mu_X$</td>
<td>$\mu_Y$</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>$\sigma_X$</td>
<td>$\sigma_Y$</td>
</tr>
</tbody>
</table>

Sample size $n_X$, Sample mean $\bar{X}$, Sample standard deviation $s_Y$.

For inference about $\mu_X - \mu_Y$, use the statistic $\sqrt{n}$, and then

- find a confidence interval for $\mu_X - \mu_Y$
- test $H_0 : \mu_X - \mu_Y = \Delta_0$ ($\Delta_0 = 0 \implies$)

To do this, we need the distribution of $\bar{X}$. Recall for independent $X$ and $Y$:

- ($\S$3.4) $\mu_{X-Y} =$
- ($\S$3.4) $\sigma^2_{X-Y} =$
- ($\S$3.4) $\sigma^2_X =$
- ($\S$4.3) If $X \sim N(\mu_X, \sigma^2_X)$ and $Y \sim N(\mu_Y, \sigma^2_Y)$, then $X - Y \sim$
- ($\S$4.8) For large $n$, the CLT says $\bar{X} \sim \approx$

It follows that, for large $n_X$ and $n_Y$, $\bar{X} - \bar{Y} \sim \approx$.
Confidence Intervals on the Difference Between Two Means

Recall that many confidence intervals have the form

(point estimate) ± (margin of error)

=(point estimate) ± (______ value for confidence) × [(estimated or true) ________________ of point estimate]

=Ĉθ ± (table value for confidence) × σĈθ

Derive a Confidence Interval

Here’s our previous derivation of a confidence interval for a normally distributed statistic:

• Consider a statistic Ĉθ as an estimator for a parameter θ, where Ĉθ ∼ N(θ, σ^2)
  (Generalize because it’s __________ to write θ than __________, Ĉθ than __________, and ____ than √(σ^2/nX + σ^2/nY).)

• Let z_α/2 = the z-score cutting off a right tail area ____ from N(0, 1) (as before), so
  P(-z_α/2 < Z < z_α/2) = ______________ (draw)

• Unstandardize using Z = ____ to get P(-z_α/2 < Ĉθ - θ/σ < z_α/2) = 1 - α; solve in two ways:
  - for Ĉθ in the middle: P(θ - z_α/2 < Ĉθ < θ + z_α/2) = 1 - α (pictured ________________)
  - for θ in the middle: P(Ĉθ - z_α/2 < θ < Ĉθ + z_α/2) = 1 - α (draw)

That is, Ĉθ ± z_α/2σ contains ____ for a proportion __________ of random samples (see picture, below). It’s the 100%(1 - α) confidence interval for θ.

The Case of a Difference of Two Means

Letting θ = ___________ and Ĉθ = ___________, gives the confidence interval we need:
Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be independent large random samples from populations with means $\mu_X$ and $\mu_Y$ and standard deviations $\sigma_X$ and $\sigma_Y$. A 100%(1 − $\alpha$) confidence interval for $\mu_X - \mu_Y$ is

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

(We usually need to use $\sigma_X \approx s_X$ and $\sigma_Y \approx s_Y$.)

e.g. (p. 273 #3) The melting points of two alloys are being compared. 35 specimens of alloy 1 melted at an average temperature of 517.0 °F, with standard deviation 2.4. 47 specimens of alloy 2 melted at an average temperature of 510.1 °F, with standard deviation 2.1. Find a 99% confidence interval for the difference between the melting points.

**Hypothesis Tests on the Difference Between Two Means**

**Recall a Test Pattern**

Many hypothesis tests ($\S$6.1, 6.3, 6.4, 7.1, 7.2, 7.3, 7.4) use test statistics of the form

$$\frac{(\text{point estimate}) - (\text{parameter value \underline{______}})}{(\text{estimated or true \underline{______}} \text{ of point estimate}}$$

This \underline{______} point estimate tells how far the estimate is from the parameter, in \underline{______}.

Here is the test for a difference of two means:

<table>
<thead>
<tr>
<th>Let $X_1, \ldots, X_{n_X}$ and $Y_1, \ldots, Y_{n_Y}$ be independent large random samples from populations with means $\mu_X$ and $\mu_Y$ and standard deviations $\sigma_X$ and $\sigma_Y$. To test $H_0 : \mu_X - \mu_Y = \Delta_0$,</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>State null and alternative hypotheses, $H_0$ and $H_1$</strong></td>
</tr>
<tr>
<td>2. <strong>Check assumptions</strong></td>
</tr>
<tr>
<td>3. <strong>Find the test statistic, $z = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$</strong> (We usually need to use $\sigma_X \approx s_X$ and $\sigma_Y \approx s_Y$.)</td>
</tr>
<tr>
<td>4. <strong>Find the $P$-value, which is an area under the $N(0,1)$ curve depending on $H_1$:</strong></td>
</tr>
<tr>
<td>$H_1 : \mu_X - \mu_Y \leq \Delta_0 \implies P$-value $= P(Z &gt; z)$, the area right of $z$</td>
</tr>
<tr>
<td>$H_1 : \mu_X - \mu_Y &lt; \Delta_0 \implies P$-value $= P(Z \leq z)$, the area left of $z$</td>
</tr>
<tr>
<td>$H_1 : \mu_X - \mu_Y \neq \Delta_0 \implies P$-value $= P(</td>
</tr>
<tr>
<td>5. <strong>Draw a conclusion</strong></td>
</tr>
</tbody>
</table>
A crayon maker is comparing the effects of two yellow dyes on crayon brittleness. Dye B is more expensive than dye A, but might produce a stronger crayon. 40 crayons are tested with each dye, and the impact strength (in joules) is measured for each. The A crayon strength averaged 2.6, with standard deviation 1.4. The B crayon strength averaged 3.8, with standard deviation 1.2.

a. Can you conclude that the mean strength of B crayons is greater than that of A crayons?

b. Can you conclude that the mean strength of B crayons exceeds that of A crayons by more than 1 J?