STT351-002
Final Exam

1-2. TREE.
20% of fields are productive.
85% of productive fields appear productive
5% of non-productive fields appear productive.
Make a tree.

1. \( P(\text{field appears productive}) \)
\[
P(\text{A}) = P(\text{PA}) + P(\text{PCA})
= 0.2(0.85) + 0.8(0.05)
\]

2. \( P(\text{field is productive | field appears productive}) \)
\[
= \frac{P(\text{PA})}{P(\text{A})}
= \frac{0.2(0.85)}{0.2(0.85) + 0.8(0.05)}
\]

3-4. CI and TEST for MEAN \( \mu \). A sample of \( n = 6 \) prescription eyeglass lenses is drawn from a process under statistical control. Each of these six is subjected to measurements which determine an overall score \( x = "\text{conformity to prescription}" \). The sample mean = 2.2 and the sample sd \( s = 2.8 \).

3. Determine the 95% CI for \( \mu \).
\[
\bar{x} \pm t_{0.05,5} \cdot \frac{s}{\sqrt{n}}
\]
\[
\bar{x} = 2.2, s = 2.8, n = 6, df = 5
\]

4. Determine the final sample size \( n_{\text{FINAL}} \) required for 95% hybrid CI
\[
\bar{x}_{\text{FINAL}} \pm 0.04
\]
\[
\frac{t_{\text{INIT}}}{\sqrt{n_{\text{FINAL}}}} = 0.04
\]
\[
\frac{2.571}{\sqrt{n_{\text{FINAL}}}} = 0.04
\]
\[
n_{\text{FINAL}} = \left( \frac{t_{\text{INIT}}}{0.04} \right)^2 = \left( \frac{2.571}{0.04} \right)^2
\]
5-6. **Probability Rules.** $P(A) = 0.6$, $P(B) = 0.9$, events $A$, $B$ are independent.

5. Complete a Venn diagram.

6. $P(B \mid \neg A) = P(B) = 0.9$

7-8. **Drawing balls.** Draws will be made without replacement and with equal probability on those remaining from $\{R R R Y Y B B B B\}$ (i.e. 3 $R$, 2 $Y$, and 4 $B$).

7. $P(R4) = \frac{1}{3}$ by what simple principle? 

   \[
P(R4) = P(R1) = \frac{3}{9} = \frac{1}{3}\]

8. Use rules of probability to PROVE $P(R2) = \frac{1}{3}$ by breaking down event $R2$ according to what happens on draw one. Cite the rules you use.

   \[
P(R2) = P(R1 \cap R2) + P(R1^c \cap R2)
   = P(R1) P(R2 \mid R1) + P(R1^c) P(R2 \mid R1^c)
   = \frac{3}{9} \cdot \frac{2}{8} + \frac{6}{9} \cdot \frac{3}{8} = \frac{29}{72} = \frac{3}{9} = \frac{1}{3}\]

9-10. **Estimates.** A sample of $n = 100$ is selected without replacement and with equal probability from a population of size 300. This sample has mean $\bar{X} = 2.1$ with $s = 0.6$.

9. Estimate the sd $\sigma_{\bar{X}}$ of sample mean $\bar{X}$

   \[
   \sqrt{\frac{n-n}{n-1}} \bar{x} = \sqrt{\frac{300-100}{300-1}} \bar{x} \cdot \frac{0.6}{\sqrt{100}}
   \]

10. Estimate the margin of error for $\bar{X}$.

   \[
   1.96 \sqrt{\frac{300-100}{300-1}} \cdot \frac{0.6}{\sqrt{100}}
   \]

   \[
   1.96 \times \frac{0.6}{10} = 0.1176
   \]
11-12. **CI for SUM of means.** A with-replacement sample of 70 parts from supplier A finds sample mean breaking strength = 560 with sample sd s = 29. Independently of this, a with-replacement sample of 90 parts from supplier B finds sample mean breaking strength 540 with sample sd s = 38.

11. Determine the 95% *z*-based CI for the difference \( \mu_A - \mu_B \).

\[
\bar{x} - \bar{y} \pm 1.96 \sqrt{ \frac{s^2_x}{n_x} + \frac{s^2_y}{n_y}}
\]

\[
(560 - 540) \pm 1.96 \sqrt{ \frac{29^2}{70} + \frac{38^2}{90}}
\]

12. Determine the 95% *z*-based CI for the **SUM** \( \mu_A + \mu_B \).

\[
\bar{x} + \bar{y} \pm 1.96 \sqrt{ \frac{s^2_x}{n_x} + \frac{s^2_y}{n_y}}
\]

\[
560 + 540 \pm 1.96 \sqrt{ \frac{29^2}{70} + \frac{38^2}{90}}
\]

13-14. **Binomial.** On average 15% of a population of vases will be damaged in shipment (i.e. \( p = 0.15 \)). Recall the discrete probability density for binomial is

\[
p(x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{for } x = 0, 1, \ldots, n.
\]

13. Determine \( p(2) \). Evaluate the binomial coefficient but do not otherwise reduce. \( n = 20 \)

\[
p(2) = \binom{20}{2} \times 0.15^2 \times 0.85^{18}
\]

\[
= \frac{20!}{2! \times 18!} \times 0.15^2 \times 0.85^{18}
\]

\[
= \frac{20 \times 19}{2} \times 0.15^2 \times 0.85^{18}
\]

14. Determine a 90% *z*-based CI for \( p \) if we find that out of a particular sample of 200 vases there are 25 damaged in shipment.

\[
\hat{p} = \frac{25}{200}
\]

\[
\hat{p} \pm 1.645 \sqrt{ \frac{\hat{p}(1-\hat{p})}{n}}
\]

\[
\frac{25}{200} \pm 1.645 \sqrt{ \frac{25 \times 0.75}{200}}
\]

\[
\frac{25}{200} \pm 1.645 \frac{\sqrt{25 \times 0.75}}{\sqrt{200}}
\]
15-16. **Poisson.** We average around 3.7 work stoppages per day. If the distribution of the number of work stoppages is Poisson recall \( p(x) = e^{-\mu} \frac{\mu^x}{x!} \), \( x = 0, 1, 2, \ldots \) ad inf.

15. In the approximation of \( P(\text{more than two stoppages tomorrow}) \) we use a standard score \( z \) of 2.5 (continuity correction of "more than two"). Determine this \( z \).

\[
\frac{z}{\sqrt{x}} = \frac{2.5 - 3.7}{\sqrt{3.7}}
\]

16. Determine \( P(\text{more than two stoppages tomorrow}) \) directly from the discrete density but express it as a finite, not infinite, sum.

\[
1 - p(0) - p(1) = 1 - e^{-3.7} \frac{3.7^0}{0!} - e^{-3.7} \frac{3.7^1}{1!}
\]

17-20. **Chi-square.** A process has been producing LEDs on average in proportions

- 25% excellent
- 10% good
- 65% avg

A random sample of 200 LEDs finds

<table>
<thead>
<tr>
<th></th>
<th>observed</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>excellent</td>
<td>48</td>
<td>50</td>
</tr>
<tr>
<td>good</td>
<td>32</td>
<td>20</td>
</tr>
<tr>
<td>avg</td>
<td>120</td>
<td>130</td>
</tr>
</tbody>
</table>

17. Fill in the expected counts above consistent with past performance.

\[
\text{eg. } 50 = \frac{25\times 200}{100}
\]

18. Determine the contribution of category "good" to the chi-square statistic.

\[
\frac{(0-50)^2}{50} = \frac{(32-20)^2}{20}
\]

19. Determine the degrees of freedom of the chi-square.

\[
Df = 2
\]

20. Determine \( p_{\text{SIG}} \) if the chi-square statistic is 11.6571.

\[
p_{\text{SIG}} = P(X^2 \geq 11.6571) \text{ between } .001 \& .005.
\]
21. **Kernel density.** Bell curves are placed at each of two points (see below). Plot the kernel density estimate. Take care to do it correctly (show five pts accurately).

![Graph of bell curves](image)

22-24. **Rules for E, Var, sd.** Random variables X, Y are independent with

\[
\begin{align*}
E X &= 6 & \text{Var } X &= 4 \\
E Y &= 9 & \text{Var } Y &= 2
\end{align*}
\]

22. \( E (XY) \overset{\text{IND}}{=} E(X) E(Y) = 6(4) \)

23. \( E (Y^2) \) (follows from \( \text{Var } Y = E (Y^2) - (E Y)^2 \))
\[
E(Y^2) = \text{Var } Y + (E Y)^2 = 2 + 9^2
\]

24. \( \text{sd } (6X - 3Y + Y - 4) \)
\[
\overset{\text{SD}}{=} \text{SD } (6X-2Y) = \sqrt{\text{Var } (6X-2Y)}
\]
\[
= \sqrt{36 \text{Var } X + 4 \text{Var } Y} = \sqrt{36(4) + 4(2)}
\]
25. **Plot regression line.** Parts are sampled with-replacement and scored \((x, y)\) where 
\(x = \) serial number of part \(y = \) hardness.

The sample data are:
- \(\bar{x} = 1343\)
- \(s_x = 433\)
- \(n = 200\) pairs \((x, y)\)
- \(\bar{y} = 12.7\)
- \(s_y = 1.1\)
- \(r = 0.7\)

What is the value \(y\) for a point on the regression line with \(x = 1343 + 433\) (i.e. one sample sd above the sample mean in the x-scale)?

\[
\text{ANS: } \bar{y} + r \left( \frac{s_y}{s_x} \right) = 12.7 + 0.7(1.1)
\]

26. **Proportionally stratified.** A population of motors is stratified by supplier

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20%</td>
</tr>
<tr>
<td>B</td>
<td>10%</td>
</tr>
<tr>
<td>C</td>
<td>70%</td>
</tr>
</tbody>
</table>

A stratified sample of motors produces the following sample means by stratum

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Sample Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2.4</td>
</tr>
<tr>
<td>B</td>
<td>2.7</td>
</tr>
<tr>
<td>C</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Estimate the population mean \(\mu\) from the above data.

\[
\bar{x} = \frac{\sum w_i \bar{x}_i}{\sum w_i} = 0.2(2.4) + 0.1(2.7) + 0.7(2.0)
\]

27. **Calculating SD.** For the following discrete distribution calculate the standard deviation \(\sigma\).

<table>
<thead>
<tr>
<th>(x)</th>
<th>(p(x))</th>
<th>(x) (p(x))</th>
<th>(x^2) (p(x))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

\[
E_x = 0.2 \quad E_{x^2} = 0.2
\]

\[
\sigma = \sqrt{\text{Var}(X)} = \sqrt{E_{x^2} - (E_x)^2} = \sqrt{0.2 - 0.2^2} = \sqrt{0.18}
\]

(usual formula)

\[\frac{U}{P(P)}\text{ FOR} \quad \text{SD-1 Scores}\]
28-29. **Multiple regression.** A random sample of 400 of our products is selected from stores nationwide. Each is scored for

- \( y \) = selling price
- \( x_1 = 1 \) if store is major retailer, 0 if not
- \( x_2 \) = quantity ordered by store

A multiple linear regression is fit to this data resulting in the fitted model

\[
y = 44.75 - 7.80 \times x_1 - 0.083 \times x_2
\]

28. Determine the average effect on price (according to the fitted model) occasioned by adding 500 to the order and switching from a major retailer to one that is not a major retailer.

\[
\Delta y = 7.8 \text{ (from } -7.801 \text{ to } -7.800\text{)}
\]

\[
\Delta y = 7.8 - 0.083 (500) \text{ (from } -0.083 x_2 \text{ to } -0.083 \text{)}
\]

\[
\Delta y = 7.8 - (0.083)(500)
\]

29. Compare the 95% CI of \( \mu \) based on \( \bar{y} = 42.76 \) with that based on the regression based estimator if

- sample multiple correlation is \( \hat{R} = 0.6 \),
- regression based estimator works out to 37.80.

95% CI using \( \bar{y} \)

\[
95\% \text{ CI using } \bar{y} = 42.76 \pm 1.96 \frac{94}{\sqrt{400}}
\]

95% CI using regression based estimator

\[
37.80 \pm 1.96 \sqrt{1 - 0.6^2} \frac{94}{\sqrt{400}}
\]

30. **t-TEST.** A process is in control. Each part produced is score \( x \) = finishing time. A sample of 12 will be used to monitor the process in a test of the null hypothesis

\( H_0: \mu_x = 5 \) (minutes) vs \( H_1: \mu_x \neq 5 \) with \( \alpha = 0.1 \)

30. If the test statistic for a sample of 12 evaluates to \( t = 2.8 \) what action is taken by the test? Indicate your reasoning.

\[
\text{DF} = n-1 = 12-1 = 11
\]

\[
\text{REJECT } H_0 \text{ IF } |t_{\text{STAT}}| > t_{\text{CRITICAL}}
\]

\[
0 \text{ IF } |t_{\text{STAT}}| > 1.796
\]

OR \[
\text{TABLE III B.S.} = P(1T1 > 2.8) = 0.009
\]

\[
\text{DF} = 11 \text{ Cumulative}
\]

\[
0.009
\]

\[
0.095
\]

\[
0.195
\]

\[
0.284
\]

\[
1.796
\]