• A general inference framework:

If $X_1, X_2, \ldots, X_n$ is SRS from some dist^n involving the parameter of interest θ, then for inference on θ we usually look at an estimator of θ, say $\hat{\theta}$, such that the distribution of

$$\frac{\hat{\theta}-\theta}{SE(\hat{\theta})} \text{ is free of } \theta \text{ and is completely known.}$$

This is called a 'pivotal quantity'.

eg. If $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$; $\theta = \mu$ and σ^2 known,

$$\frac{\hat{\theta}-\theta}{SE(\hat{\theta})} = \frac{\bar{X}-\mu}{SE(\bar{X})} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \left(\text{as } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)\right)$$

which forms the basis of Z-based CI's and tests.

In general, with a predetermined small α,

$100(1-\alpha)\%$ CI for θ : $\hat{\theta} \pm CV_{\alpha/2} \cdot SE(\hat{\theta})$

" UCB for θ : $(-\infty, \hat{\theta} + CV_\alpha \cdot SE(\hat{\theta}))$

" LCB for θ : $(\hat{\theta} - CV_\alpha \cdot SE(\hat{\theta}), \infty)$

Where CV_α is the upper α-quantile or the appropriate 'critical value' from the dist^n of $(\hat{\theta}-\theta)/SE(\hat{\theta})$.

Also remember that this may be an approximate result (often based on CLT), making the CI's and tests approximate, which means you cannot be entirely sure about the 'coverage probability' being $(1-\alpha)$ or that $P(\text{Type-I error}) = \alpha$.

This dist^n is usually $N(0,1)$ or $T_{(n-1)}$ and as a result, you have to get CV_α $(= z_{1-\alpha}$ or $t_{\alpha;n-1}$ respectively) from the appropriate table.

Also remember that if $SE(\hat{\theta})$ involves θ or some other unknown parameter, you will have to replace $SE(\hat{\theta})$ by $\hat{SE}(\hat{\theta})$, often using the sample SD S or a simple plug-in estimator. This may change the distribution of the pivotal quantity.

eg. if σ^2 is unknown, $SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ is replaced by $\hat{SE}(\bar{X}) = \frac{S}{\sqrt{n}}$ which changes the dist^n from $N(0,1)$ to $T_{(n-1)}$.

Now, the standardized 'Test Statistic' (TS) becomes

$$TS = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})} \text{ or } \frac{\hat{\theta} - \theta_0}{\hat{SE}(\hat{\theta})}$$, and the rejection regions become $\{TS > CV_\alpha\}$, $\{TS < -CV_\alpha\}$ and $\{|TS| > CV_{\alpha/2}\}$ for the level-α upper tailed, lower tailed and two tailed tests respectively. (Assuming symmetry of the concerned dist^n) $[H_0: \theta = \theta_0$ vs. $H_a: \theta \underset{\text{any one of these}}{\gtrless} \theta_0]$

Now let us look at a different example.

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Ber(p)$; $\theta = p \in (0,1)$

Inference of p is based on the dist^n of $X = \sum_{i=1}^{n} X_i$, which is $Bin(n,p)$. For small n, an exact test / CI based on the discrete $Bin(n,p)$ dist^n can be developed, but we will not discuss that. (This is discussed in P349 of the textbook (ed.9))

We will assume a large n, so that we can use CLT. (i.e. $np \geq 10$, $n(1-p) \geq 10$, or just assume that the dist^n of X is sufficiently symmetric)

In that case, $X \overset{\text{approx.}}{\sim} N(np, np(1-p))$

equivalently, $\hat{p} = \frac{X}{n} \overset{\text{approx.}}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$

This gives us: $\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\text{approx.}}{\sim} N(0,1)$

which is exactly like the general framework discussed above. The only problem is that $SE(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$ involves the unknown p and needs to estimated. So we replace it by $\hat{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ and assume that $\hat{SE}(\hat{p}) \approx SE(\hat{p})$ for the large n.

Our results were approximate anyway and this just has an extra level of approximation.

Now, the $100(1-\alpha)\%$ approx. CI, UCB and LCB for p are:

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \; ; \; \left(-\infty, \hat{p} + z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

and $\left(\hat{p} - z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \infty\right)$ respectively.

Similarly the test statistic: $TS = \dfrac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ can be used to test $H_0: p = p_0$ vs. $H_a: p \gtrless p_0$ (approximately) with the rejection regions $\{TS > z_{1-\alpha}\}$, $\{TS < -z_{1-\alpha}\}$ and $\{|TS| > z_{1-\frac{\alpha}{2}}\}$ for the upper, lower and two-tailed alternative hypotheses respectively.

Note: Your book has a much more detailed discussion on population proportion p, but you don't need these for the exam

In fact, for the hypothesis tests, you don't need the extra approximation through $\widehat{SE}(\hat{p})$.

Since, $\dfrac{\hat{p} - p}{SE(\hat{p})} \sim N(0,1)$ approx., we can say $\dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$ (approx.) under the null hypothesis $H_0: p = p_0$.

Since we develop the TS and rejection regions based on H_0 (with the necessity: $P(\text{Type-I error}) = \alpha$), we can simply use the Test statistic:

$$TS = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad \text{instead of} \quad \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}.$$

The first one is slightly better in terms of approximation, but should be okay for a large n. The rejection regions remain unchanged.

You can use either one in the exam in a similar situation. Also, the $TS = (\hat{p} - p_0)/\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ shares a direct connection with the confidence regions for p developed earlier that the slightly better one, i.e. $TS = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ does not.

Note that by a similar logic, for part e of Q3 in the practice exam, $Z = \frac{\bar{X} - \lambda_0}{SE(\bar{X})} = \frac{\hat{\lambda} - \lambda_0}{\sqrt{\lambda_0/n}} = \frac{5.75 - 5}{\sqrt{5/45}} = 2.25$ provides a slightly better test statistic. With the same rejection: $\{|z| > 2.576\}$, you arrive at the same conclusion. But you will get full credit with the version in the solution.

- **Prediction Interval:**

Let $X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$; σ^2 is known.

Instead of μ, we will infer on X_{n+1}, a single future observation from the same dist^n.

Firstly, the usual estimator for X_{n+1} from our data clearly is $\bar{X}$ (You may say: $\hat{X}_{n+1} = \bar{X}$)

Now let's look at $\bar{X} - X_{n+1} = \frac{1}{n}(X_1) + \dots + \frac{1}{n}(X_n) + (-1)(X_{n+1})$ which is a linear combination of $X_1, X_2, \dots, X_{n+1}$. (iid)

From section 5.5 results,

$$E(\bar{X} - X_{n+1}) = \frac{1}{n}(\mu) + \dots + \frac{1}{n}(\mu) + (-1)(\mu) = 0,$$

$$V(\bar{X} - X_{n+1}) = \left(\frac{1}{n}\right)^2\sigma^2 + \dots + \left(\frac{1}{n}\right)^2\sigma^2 + (-1)^2 \cdot \sigma^2 = \frac{n}{n^2}\sigma^2 + \sigma^2 = \sigma^2\left(1 + \frac{1}{n}\right)$$

and $\bar{X} - X_{n+1} \sim N\left(0, \sigma^2\left(\frac{n+1}{n}\right)\right)$

$$\Leftrightarrow \frac{\bar{X} - X_{n+1}}{\sigma\sqrt{\frac{n+1}{n}}} \sim N(0,1),$$ which is exactly similar to the general framework (replace θ by X_{n+1}, $\hat{\theta}$ by $\bar{X}$ and $SE(\hat{\theta})$ by $\sigma\sqrt{\frac{n+1}{n}}$.)

Now, using the general logic,

$100(1-\alpha)\%$ CI for X_{n+1} is: $\bar{X} \pm z_{1-\frac{\alpha}{2}} \cdot \sigma\sqrt{\frac{n+1}{n}}$

This is called the $100(1-\alpha)\%$ Prediction interval (PI) as X_{n+1} is an unobserved future data

Similarly the $100(1-\alpha)\%$ prediction bounds are:

$\left(-\infty, \bar{X} + z_{1-\alpha} \cdot \sigma\sqrt{\frac{n+1}{n}}\right)$ (upper) and

$\left(\bar{X} - z_{1-\alpha}\sigma\sqrt{\frac{n+1}{n}}, \infty\right)$ (lower)

You can develop tests for X_{n+1} in the same way.

If σ^2 is unknown, you replace it by S (calculated from $X_1, X_2, \ldots, X_n$) and developed t-based prediction intervals and bounds in the same way.

e.g. $100(1-\alpha)\%$ PI will be: $\bar{X} \pm t_{\alpha/2;\, n-1} \cdot S\sqrt{\frac{n+1}{n}}$.

One important aspect of PI's is that they are always wider than the corresponding CI's for μ (with the same α). That's because $\sigma\sqrt{\frac{n+1}{n}} > \sigma/\sqrt{n}$.

When we infer on μ, we have to realize that μ itself is non-random (or, fixed) and the only uncertainty in its inference comes from its estimator $\bar{X}$ (i.e. $V(\bar{X}) = \sigma^2/n$). On the other hand, X_{n+1} is itself random and its uncertainty is added to the uncertainty of its estimator $\bar{X}$ (i.e. $V(X_{n+1}) + V(\bar{X}) = \sigma^2\left(\frac{n+1}{n}\right) > V(\bar{X}) = \frac{\sigma^2}{n}$.)

So PI's are wider than CI's to account for this extra uncertainty.

- **Minimum Sample size calculation for t based CI's:**

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$; $\theta = \mu$ and σ^2 is unknown

What is the minimum sample size (n_{min}) so that the length of the $100(1-\alpha)\%$ CI for μ is $\leq l_{max}$.

i.e. $\left(\frac{2S\, t_{\alpha/2;\, n-1}}{\sqrt{n}}\right) \leq l_{max}$

$$\Leftrightarrow \quad n \geq \left(\frac{2S\, t_{\alpha/2;\, n-1}}{l_{max}}\right)^2$$

This means: $n_{min} = \left\lceil \left(\frac{2S\, t_{\alpha/2;\, n_{min}-1}}{l_{max}}\right)^2 \right\rceil$ which is circular

The problem with this expression is that the expression for n involves n itself*. This problem arises because to calculate n_{min}, you need to guess the unknown σ by S which itself needs a sample.

* through not only $t_{\alpha/2;\, n-1}$, but also S which is based on a sample of size n.

So we will need a preliminary study (or a pilot study) to first estimate σ by S from that pilot data. Let's say the sample size for that study is n' and the sample SD from that sample is S'.

Now, we can say: $n_{min} = \left\lceil \left(\frac{2S'\, t_{\alpha/2;\, n'-1}}{l_{max}}\right)^2 \right\rceil$.

(approximately)

This formula works because n_{min} is very likely to be $> n'$. Otherwise your preliminary study would already give you $l \leq l_{max}$. But usually the preliminary study is done with a small sample size n'.

Since $n_{min} > n'$; $t_{\alpha/2;\, n_{min}-1} < t_{\alpha/2;\, n'-1}$.

Hence with this formula you are getting a larger n_{min} than what is truely required. But this is still okay as you are making sure that: $l \leq l_{max}$.

Moreover in any given problem, you have one sample completely known. You can always use that as your preliminary sample, especially if that sample has length larger than your required l_{max}.