Ex 6: A group of 40 men and 40 women are polled on how many hours of TV they watch per week. Group A (men): avg 10.2 hrs, std dev = 3.1 hrs
Group B (women): avg 7.5 hrs, std dev = 4.6 hrs

a.) Find a confidence interval for the true difference in mean hours of TV watched between men & women with a confidence coefficient of 1-α = 0.90.

\[ z_{0.05} = 1.645 \]
\[ \theta_{u} = \theta + z_{(0.05/2)} \sigma_{\theta} \]
\[ \theta_{l} = \theta - z_{(0.05/2)} \sigma_{\theta} \]

So \[ \bar{y}_{1} - \bar{y}_{2} \pm 1.645 \sqrt{\frac{3.1^2}{40} + \frac{4.6^2}{40}} \]
\[ = 2.7 \pm 1.645(0.8771) \rightarrow [1.257, 4.143] \]

b.) How confident are we that men watch an average of ≥ 2 more hours of TV than women?

\[ M_{1} - M_{2} ≥ 2 \]?

\[ \frac{P(\hat{\theta}_{1} \leq \theta)}{\bar{y}_{1} - \bar{y}_{2}} = \frac{\hat{\theta}_{u} - \theta}{\sigma_{\theta}} \]
\[ \frac{2 = \bar{y}_{1} - \bar{y}_{2} - 2(-0.7981)}{\sqrt{\frac{3.1^2}{40} + \frac{4.6^2}{40}}} \]
\[ 2 = 2.7 - 2(0.8771) \]
\[ \rightarrow z_{(0.05)} = 0.7981 \]

\[ \alpha = P(Z < -0.7981) = P(Z > 0.7981) = 0.2119 \]

We want \[ 1 - \alpha \]
\[ 1 - (0.2119) = 0.7881 \]

So we are 78.81% confident that men watch at least 2 more hours of TV per week than women.
7.2 Sampling Distributions

Already learned about normal distribution & Central Limit Theorem (CLT). Now learn about 3 other distributions that will come up soon: χ², T, & F.

From Sec 4.6, learned about gamma distribution:

Defn: A RV Y has a gamma distribution with parameters α ≥ 0 & β > 0 if

\[ f(y) = \begin{cases} \frac{y^{α-1} e^{-y/β}}{β^α Γ(α)} & 0 ≤ y < ∞ \\ 0 & \text{elsewhere} \end{cases} \]

where \( Γ(α) = \int_0^∞ y^{α-1} e^{-y} dy \)

Defn: A RV Y has a chi-square distribution with n degrees of freedom if \( α = \frac{n}{2} \) & β = 2.

So \( f(y) = \begin{cases} \frac{y^{n/2-1} e^{-y/2}}{2^{n/2} Γ(n/2)} & 0 ≤ y < ∞ \\ 0 & \text{elsewhere} \end{cases} \)

Don't need to know this for Math 123.

Thus: let \( Y_1, \ldots, Y_n \) be independent, normal RVs w/ mean μ & variance σ². Then \( Z_i = \frac{Y_i - μ}{σ} \) are independent, standard normal RVs &

\[ \sum_{i=1}^n Z_i^2 \text{ has a } χ² \text{ distribution with } n \text{ degrees of freedom.} \]

What does χ² distribution look like for various values of n?

1.) n=1

2.) n=2

3.) n=3

Note: χ² has a local max at \( y = n-2 \) if \( n > 2 \)

4.) n=10

Table 6, Appendix 3 (pg 50 - 857) gives \( P(χ² > χ²_{α, n}) = α \)

Note: This is opposite of Z tables! Z tables give probs for various y's, χ² table gives y's for various probs!
For $\chi^2$ with $n$-degrees of freedom: $E(\chi^2) = n$, $\nu(\chi^2) = 2n \leq n\gamma(\alpha/2)$ in HW.

Now look at relationship between $S^2$ and $\chi^2$:

Recall $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$ is an unbiased estimator for $\sigma^2$.

Thm: Let $Y_1, \ldots, Y_n$ be independent normal RVs with mean $\mu$ and variance $\sigma^2$. Then $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$ has a $\chi^2$ distribution with $n-1$ degrees of freedom and $\bar{Y}$ and $S^2$ are independent.

Proof idea: $\chi^2 = \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left( \frac{Y_i - \mu}{\sigma} \right)^2$ has $n$ degrees of freedom. Here, $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} \left( \frac{Y_i - \bar{Y}}{\sigma} \right)^2$ is almost like $\sum_{i=1}^{n} Z_i^2$, but using $\bar{Y}$ instead of $\mu$.

So instead of $n$ degrees of freedom, have $n-1$ degrees of freedom.

Notation: Use $X_n^2$ to denote $\chi^2$ with $n$ degrees of freedom.

So why is this theorem useful?

So far: $\sigma^2 = \text{population variance (true value of parameter)}$.

1. We know $\sigma^2$, so plug in for error bound.
2. Don't know $\sigma^2$, so plug in estimate for $\sigma^2$ or "sample variance" for error bound.

EX: $\frac{\sigma}{\sqrt{n}}$, $\sqrt{\frac{\sigma^2 + \hat{\sigma}^2}{n}} \leq \text{plugged in values for } \sigma^2, \hat{\sigma}^2, \sigma_{x-y}^2$

Why can we do this? Full answer in Sec 9.3
Partial answer: For large $n$, sample variance $\approx \sigma^2$.

Q: What if $n$ is not large?

Then can't use $Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$ anymore, since sample variance $\neq \sigma^2$.

A: Define new RV: $\frac{\bar{Y} - \mu}{S/\sqrt{n}}$.

So $S$ = sample variance, $Y_i$ are normal.

$\frac{\bar{Y} - \mu}{S/\sqrt{n}} = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} = \frac{Z}{S/\sigma} = \frac{Z}{\sqrt{\frac{X_n^2}{n-1}}} \leq \text{call this } T_{n-1}$.
If $Z$ & $X_n^2$ are independent, then

$$T_n = Z / \sqrt{X_n^2 / n}$$

has a "t distribution" with $n$ degrees of freedom.

For a t-distribution with $n$ degrees of freedom, $E(T_n) = 0$, $V(T_n) = \frac{n}{n-2}$ if $n > 2$.

Note: as $n \to \infty$, $V(T_n) \to 1$.

What does $T_n$ look like?

Fatter $Z$:

In Sec 9.3, show $T_n \to Z$ as $n \to \infty$.

Another useful distribution:

**Defn:** Let $X_n^2$ & $X_n^2$ be independent. Then $F_{n_2}^{n_1} = \frac{X_n^2 / n_1}{X_n^2 / n_2}$ has a $F$ distribution with $n_1$ numerator df & $n_2$ denominator df.

For $F_{n_2}^{n_1}$, $E(F_{n_2}^{n_1}) = \frac{n_2}{n_2 - 2}$ if $n_2 > 2$ & $V(F_{n_2}^{n_1})$ is a complicated rational function if $n_2 > 4$.

What does $F_{n_2}^{n_1}$ look like?

Usually looks something like this.

Note: In textbook appendices, have tables for both $T_n$ & $F_{n_2}^{n_1}$ and tables are just like $X_n^2$ (opposite of $Z$ tables).

8.8 Small Sample Confidence Interval for $\mu$

So far, $n$ large, can use sample variance to approx population variance $\sigma^2$

So $\frac{\bar{Y} - \mu}{S/\sqrt{n}} \to \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$ when $n$ is large (proof in 9.3).

What if $n$ is small? Can't use $Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$.

Instead use t-distribution $\frac{\bar{Y} - \mu}{S/\sqrt{n}} = T_{n-1}$ (Note: T-distribution assumes $Y_i$ are normal).

Confidence interval: $x_{a/2}$

$$P(-t_{n-1} \leq T_{n-1} \leq t_{n-1}) = 1 - \alpha$$

Just like $Z$, since symmetric, so two-sided

Confidence interval for $\mu$:

$$\bar{Y} \pm t_{n-1} \frac{S}{\sqrt{n}}$$

One-sided: $\bar{Y} - t_{n-1} \frac{S}{\sqrt{n}}$ is lower bound & $\bar{Y} + t_{n-1} \frac{S}{\sqrt{n}}$ is upper bound.

**Important Note:** Can use T-dist to estimate $\mu$, cannot use to estimate $\hat{p} = \frac{Y}{n}$. 

9
8.9 Confidence Intervals for $\sigma^2$

So far, we’ve used sample variance to estimate $\sigma^2$ when we were trying to estimate $\theta = \bar{Y}, \bar{Y}, \bar{Y}_1 - \bar{Y}_2, \bar{Y}_1 - \bar{Y}_2$ for large $n$ (use $Z$) and $\theta = \bar{Y}, \bar{Y}_1 - \bar{Y}_2$ for small $n$ (use $T$)

What if we want a confidence interval for $\sigma^2$?

Recall: $\frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \frac{(n-1)S^2}{\sigma^2}$ has a $\chi^2_{n-1}$ distribution if $Y_i$ are normal

$\implies P \left( \chi^2_{1 - \frac{\alpha}{2}} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha} \right) = 1 - \alpha$ 

$\implies P \left( \chi^2_{1 - \frac{\alpha}{2}} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2} \right) = 1 - \alpha$

So $P \left( \frac{(n-1)S^2}{\chi^2_{1 - \frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha/2}} \right) = 1 - \alpha$

Q: Why did we find confidence interval this way, instead of like before: $P \left( |S^2 - \sigma^2| \leq b \right)$

A: $S^2$ is not normal or $t$! So no point in expressing $b$ in terms of $\sigma^2$

Ex 8: Let’s say I home brew beer and produce an inconsistent batch of IPAs. The amount of ABV varies in 5 batches: 5.4, 5.9, 6.3, 6.7, 7.0.

Estimate $\sigma^2$ with a confidence coefficient of $1 - \alpha = 0.90$.

$\bar{Y} = \frac{5.4 + 5.9 + 6.3 + 6.7 + 7.0}{5} = 6.26$

$S^2 = \frac{1}{4} \left( (5.4 - 6.26)^2 + (5.9 - 6.26)^2 + \cdots + (7.0 - 6.26)^2 \right) = 0.403$

$\frac{\alpha/2 = 0.05}{X^2_{1 - \frac{\alpha}{2}}} = X^2_{0.05} = \chi^2_{0.05} = 9.4877$

$X^2_{\alpha/2} = X^2_{0.95} = \chi^2_{0.95} = 0.7107$

$\left( \frac{4 \cdot (0.403)}{9.4877}, \frac{4 \cdot (0.403)}{0.7107} \right) = (0.1699, 2.2682)$ is 90% conf int for $\sigma^2$

This implies $(0.4123, 1.5060)$ is a 90% conf int for $\sigma$, where $s = \sqrt{0.403} = 0.6348$
9.1 9.2 Intro/Efficiency

So far we’ve used point estimators $\overline{Y}$, $\hat{p} = \frac{Y}{n}$, $\hat{P} - \overline{Y}$, $\hat{P} - \overline{P}$, $S^2$.

Q: Why did we choose these? Are there others that are “better”?

Ex: We saw $S^2$ was biased, so prefer to use unbiased $S^2$.

So we want our point estimators to be unbiased, but are these other qualities we would like the point estimators to have? How to choose the best one?

One property we’d like to have: relative efficiency.

Defn Given two unbiased estimators $\hat{\theta}_1$ & $\hat{\theta}_2$ of a parameter $\theta$, with variances $\text{var}(\hat{\theta}_1)$ & $\text{var}(\hat{\theta}_2)$, respectively, then the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$, is

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{var}(\hat{\theta}_2)}{\text{var}(\hat{\theta}_1)}$$

So if $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) > 1 \rightarrow \text{var}(\hat{\theta}_2) > \text{var}(\hat{\theta}_1)$, so $\hat{\theta}_1$ is “relatively more efficient”.

if $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) < 1 \rightarrow \text{var}(\hat{\theta}_2) < \text{var}(\hat{\theta}_1)$, so $\hat{\theta}_2$ is “relatively more efficient.”

9.3 Consistency.

Let’s say we have a coin, and we don’t know if it’s a fair coin, so we don’t know $p = \text{prob of heads}$.

If we want to determine $p$, can flip it many times and calculate $\frac{Y}{n}$, where $n = \# \text{trials}$, $Y = \# \text{number of heads}$.

What happens as $n \to \infty$?

$\frac{Y}{n} \to$ closer & closer to true value, i.e. it “converges in probability to $p$.”

Rigorous: Let $\epsilon > 0$. Then $\lim_{n \to \infty} \text{P}(1 \frac{Y}{n} - p | \leq \epsilon) = 1$

Defn: The estimator $\hat{\theta}$ is a consistent estimator of $\theta$ if, for any positive number $\epsilon$, $\lim_{n \to \infty} \text{P}(|\hat{\theta} - \theta| \leq \epsilon) = 1$

Or equivalently, $\lim_{n \to \infty} \text{P}(|\hat{\theta} - \theta| > \epsilon) = 0$

Note: $\hat{\theta}$ is a function of $n$.

Ex: $\hat{\theta} = \overline{Y} = \frac{Y_1 + \ldots + Y_n}{n}$ or $\hat{p} = \frac{Y}{n}$ or $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$

Book uses notation $\hat{\theta}_n$ to clarify this.

Thm: An unbiased estimator $\hat{\theta}$ for $\theta$ is a consistent estimator of $\theta$ if $\lim_{n \to \infty} \text{var}(\hat{\theta}) = 0$

(Converse is not necessarily true).
Can use this to prove Y is consistent!

Ex 9: Let \(Y_1, \ldots, Y_n\) be a random sample. Show that \(\overline{Y}\) is a consistent estimator of \(\mu\).

Note: \(E(\overline{Y}) = \mu\), \(V(\overline{Y}) = \frac{\sigma^2}{n}\). So since \(V(\overline{Y}) \to 0\) as \(n \to \infty\), then \(\overline{Y}\) is a consistent estimator of \(\mu\).

This is called Weak Law of large numbers: \(\lim_{n \to \infty} P(\overline{Y} - \mu \leq \varepsilon) = 1\).

Following theorem helps us prove consistency for functions of estimators:

**Thm** Suppose \(\hat{\theta}\) converges in probability to \(\theta\) and that \(\theta'\) converges in probability to \(\theta'\).

Then:

a.) \(\hat{\theta} + \theta'\) conv in prob to \(\theta + \theta'\)

b.) \(\hat{\theta} \cdot \theta'\) conv in prob to \(\theta \cdot \theta'\)

c.) \(\frac{\hat{\theta}}{\theta'}\) conv in prob to \(\frac{\theta}{\theta'}\) if \(\theta' \neq 0\)

d.) If \(g\) is continuous at \(\theta\), then \(g(\hat{\theta})\) conv in prob to \(g(\theta)\)

Ex 10: Let \(Y_1, \ldots, Y_n\) be a random sample. Show that \(S^2\) is a consistent estimator of \(\sigma^2 = V(Y_i)\).

Note: \(S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \frac{1}{n-1} \sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\)

(This problem does not say \(Y_1, \ldots, Y_n\) from normal distribution, so can't use \(\chi^2_{n-1}\)).

Try to use Weak law of large numbers:

\(\frac{n}{n-1} \left( \frac{1}{n} \sum_{i=1}^{n} Y_i^2 \right) = \frac{n}{n-1} (\overline{Y}^2)\)

This is average of \(Y_i^2\), where \(E(Y_i^2) = \mu_2\), \(V(Y_i^2) = E(Y_i^4) - (E(Y_i^2))^2 = \mu_4 - (\mu_2)^2 < \infty\)

So \(\frac{n}{n-1} \sum_{i=1}^{n} Y_i^2\) conv in prob to \(\mu_2\).

Also, \(\overline{Y}\) conv in prob to \(\mu\), so \(\overline{Y}^2\) conv in prob to \(\mu^2\) by part (d.) of Thm.

So since \(\lim_{n \to \infty} \frac{n}{n-1} = 1\), then full expression conv in prob to \(\mu_2 - \mu^2 = E(Y_i^2) - (E(Y_i))^2 = V(Y_i) = \sigma^2\).
One final theorem necessary to show that \[ \frac{\overline{Y} - \mu}{S/\sqrt{n}} \to Z \]

Thm: Suppose \( U_n \) has a distribution \( F_n \) that converges to \( Z \) as \( n \to \infty \).

If \( W_n \) conv in prob to 1, then \( U_n/W_n \) converges to \( Z \).

Ex 11: Suppose \( Y_1, \ldots, Y_n \) is a random sample with mean \( \mu \) and variance \( \sigma^2 \).

Show that \[ \frac{\overline{Y} - \mu}{S/\sqrt{n}} \to Z. \]

In Ex 10, showed that \( S^2 \) conv in prob to \( \sigma^2 \).

So \( S \) conv in prob to \( \sigma \).

So \( S/\sigma \) conv in prob to 1.

Central Limit Thm says \[ \frac{\overline{Y} - \mu}{S/\sqrt{n}} \to Z \] as \( n \to \infty \).

So \( U_n/W_n = \frac{\overline{Y} - \mu}{S/\sqrt{n}} / S/\sigma \to Z \)

So this is why we were able to plug in sample variance for \( \sigma^2 \) for large \( n \).

Also, this says if \( Y_i \) are normal, then \[ \frac{\overline{Y} - \mu}{S/\sqrt{n}} = T_n \text{ converges to } Z \quad \text{as } n \to \infty \]

\[ T_n \to Z \]