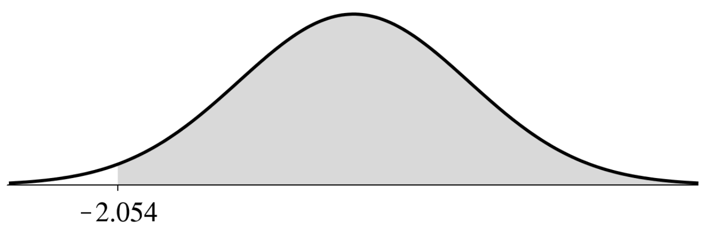
Math 224 **Fall 2021**
Homework 6 **Drew Armstrong**

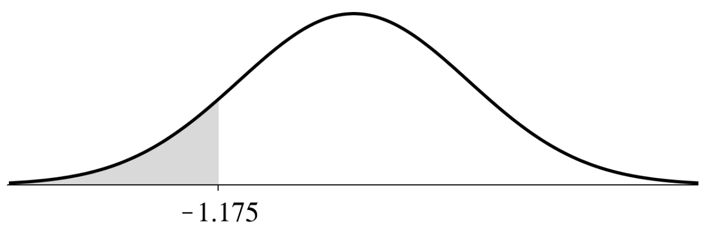
1. Let $Z \sim N(0,1)$. Use the attached tables to solve for a.

(a) $P(Z > a) = 98\%$
(b) $P(Z < a) = 12\%$
(c) $P(|Z| > a) = 5\%$
(d) $P(|Z| < a) = 50\%$

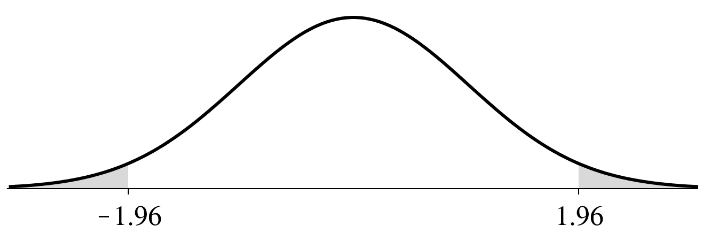
(a): There is 98% to the right of -2.054:



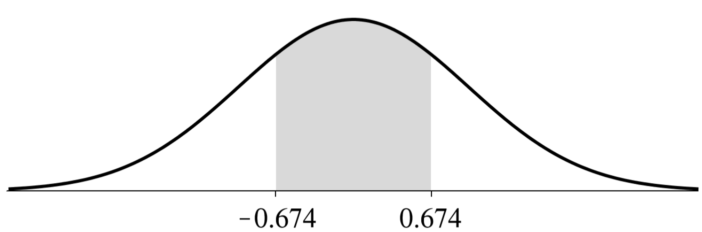
(b): There is 12% to the left of -1.175:



(c): There is 5% where $Z < -1.96$ or $Z > 1.96$ (i.e., there is 2.5% in each tail):



(d): There is 50% where $-0.674 < Z < 0.674$:

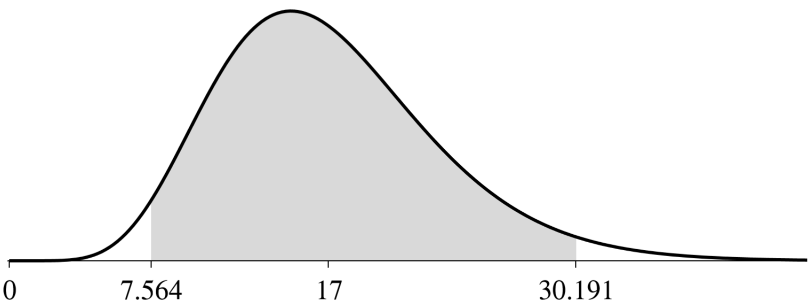


2. Let $Q \sim \chi^2(17)$. Use the attached tables to solve the following.

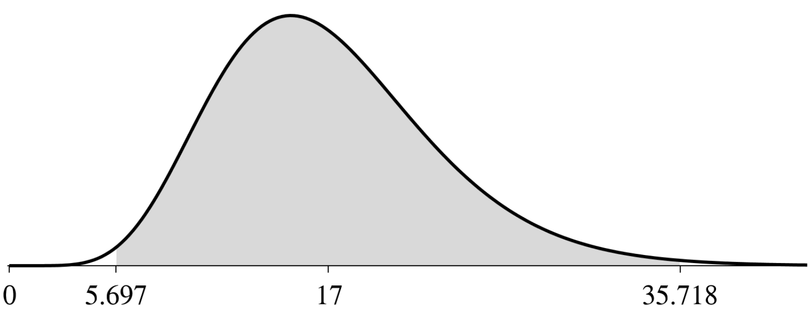
(a) Find numbers a, b such that $P(a < Q < b) = 95\%$.
(b) Find numbers a, b such that $P(a < Q < b) = 99\%$.

Remark: These problems have infinitely many possible answers, but only one answer can be found using the attached table.

(a): If we assume that each tail has the same probability 2.75% then $a = \chi^2_{97.5\%}(17) = 7.564$ and $b = \chi^2_{2.75\%}(17) = 30.191$:



(b): If we assume that each tail has the same probability 0.5% then $a = \chi^2_{99.5\%}(17) = 5.695$ and $b = \chi^2_{0.5\%}(17) = 35.718$:



3. Let $X_1, \ldots, X_n$ be an iid sample from a normal distribution $N(\mu, \sigma^2)$. If μ is **known** and σ^2 is **unknown** then we will consider the estimator

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2.$$

Show that $E[\hat{\sigma}^2] = \sigma^2$. [Hint: First show that $E[X_i^2] = \mu^2 + \sigma^2$. Then use linearity.]

Since $X_i \sim N(\mu, \sigma^2)$ for each i we have $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$, hence

$$\begin{aligned} E[X_i^2] - E[X_i]^2 &= \text{Var}(X_i) \\ E[X_i^2] - \mu^2 &= \sigma^2 \\ E[X_i^2] &= \mu^2 + \sigma^2. \end{aligned}$$

Using this fact, we compute the expected value of $\hat{\sigma}^2$:

$$\begin{aligned}
E\left[\hat{\sigma}^2\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\mu)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}E\left[(X_i-\mu)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}E\left[X_i^2-2\mu X_i+\mu^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(E\left[X_i^2\right]-2\mu E[X_i]+\mu^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(E\left[X_i^2\right]-2\mu\mu+\mu^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\mu^2+\sigma^2-2\mu^2+\mu^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2\right) \\
&= \frac{1}{n}\cdot n\sigma^2 \\
&= \sigma^2.
\end{aligned}$$

In other words, if the population mean μ is known then we can use $\hat{\sigma}^2$ as an unbiased estimator for the unknown population variance σ^2. This situation rarely arises in practice.

4. Let $Z_1, \ldots, Z_n$ be an iid sample from a standard normal distribution $N(0,1)$ and define

$$Q = Z_1^2 + Z_2^2 + \cdots + Z_n^2.$$

(a) Compute $E[Q]$.
(b) Assuming that $Z \sim N(0,1)$ implies $E[Z^4] = 3$, compute $\text{Var}(Q)$.

(a): Since $Z_i \sim N(0,1)$ we know that $E[Z] = 0$ and $\text{Var}(Z) = 1$, hence $E[Z^2] = \text{Var}(Z) - E[Z]^2 = 1$. It follows that

$$\begin{aligned}
E[Q] &= E[Z_1^2 + Z_2^2 + \cdots + Z_n^2] \\
&= E[Z_1^2] + E[Z_2^2] + \cdots + E[Z_n^2] \\
&= 1 + 1 + \cdots + 1 \\
&= n.
\end{aligned}$$

(b): If $Z \sim N(0,1)$ then we will just assume that $E[Z^4] = 3$, which is actually quite tricky to prove. Then from the remarks in part (a) we have

$$\begin{aligned}\text{Var}(Z^2) &= E[(Z^2)^2] - E[Z^2]^2 \\ &= E[Z^4] - E[Z^2]^2 \\ &= 3 - 1^2 \\ &= 2.\end{aligned}$$

Finally, since the random variables Z_i are independent, we have

$$\begin{aligned}\text{Var}(Q) &= \text{Var}(Z_1^2 + Z_2^2 + \cdots + Z_n^2) \\ &= \text{Var}(Z_1^2) + \text{Var}(Z_2^2) + \cdots + \text{Var}(Z_n^2) \\ &= 2 + 2 + \cdots + 2 \\ &= 2n.\end{aligned}$$

In other words, we have shown that a chi-squared random variable with n degrees of freedom has mean n and variance $2n$.

5. Let p be the unknown probability of heads for a certain coin. Before performing any experiments we assume that $H_0 =$ "$p = 1/2$" is true. If you flip the coin 100 times and let Y be the number of heads, what values of Y will cause you to reject H_0 in favor of $H_1 =$ "$p \neq 1/2$" at the 99% level of confidence?

We will use the sample proportion $\hat{p} = Y/n$ to estimate the unknown probability p. When testing $H_0 =$ "$p = p_0$" against $H_1 =$ "$p \neq p_0$" the rejection region is

$$|\hat{p} - p_0| > z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

In our case we have $n = 100$, $p_0 = 1/2$ and $\hat{p} = Y/100$. At the 99% level of confidence, or $\alpha = 1\%$ of significance, we have $z_{\alpha/2} = z_{0.5\%} = 2.576$, so the rejection region becomes

$$\begin{aligned}\left|\frac{Y}{100} - \frac{1}{2}\right| &> 2.576\sqrt{\frac{(1/2)(1-1/2)}{100}} \\ \left|\frac{Y}{100} - \frac{1}{2}\right| &> 2.576\sqrt{\frac{1}{400}} \\ \left|\frac{Y}{100} - \frac{1}{2}\right| &> \frac{2.576}{20} \\ |Y - 50| &> \frac{2.576 \cdot 100}{20} \\ |Y - 50| &> 12.879.\end{aligned}$$

Therefore we should reject "$p = 1/2$" in favor of "$p \neq 1/2$" when $Y > 50 + 12.879$ or $Y < 50 - 12.879$. Since Y is a whole number this becomes $Y \geq 63$ or $Y \leq 37$.

6. Let p be the unknown proportion of Americans who like broccoli. Suppose that we take a poll of $n = 1000$ Americans and $Y = 300$ tell us that they like broccoli. Use this information to compute $(1-\alpha)100\%$ confidence intervals for p when $\alpha = 5\%$, 2.5% and 1%.

We will use the sample proportion $\hat{p} = Y/n$ to estimate the unknown population proportion p. The general two-sided $(1-\alpha)100\%$ confidence interval is[1]

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

which we prefer to write as

$$p = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

In our case we have $n = 1000$, $Y = 300$ and $\hat{p} = 0.3$, so the confidence interval becomes

$$\begin{aligned} p &= 0.3 \pm z_{\alpha/2}\sqrt{\frac{0.3(1-0.3)}{1000}} \\ &= 0.3 \pm z_{\alpha/2} \cdot 0.0145. \end{aligned}$$

For $\alpha = 5\%$, 2.5% and 1% we have

$$\begin{aligned} z_{\alpha/2} &= 1.96, \\ z_{\alpha/2} &= 2.24, \\ z_{\alpha/2} &= 3.29, \end{aligned}$$

respectively. Thus we obtain the confidence intervals

$$\begin{aligned} p &= 0.3 \pm 0.0284, \\ p &= 0.3 \pm 0.0325, \\ p &= 0.3 \pm 0.0477. \end{aligned}$$

In more friendly language:

$$\begin{aligned} p &= 30\% \pm 2.84\%, \\ p &= 30\% \pm 3.25\%, \\ p &= 30\% \pm 4.77\%. \end{aligned}$$

7. Assume that the weight of pumpkins grown on a certain farm is $N(\mu, \sigma^2)$. In order to estimate μ and σ^2 we weighed a random sample of 7 pumpkins (in pounds):

10.7	8.5	9.1	10.3	13.7	9.7	9.3

Use the attached tables to find 95% confidence intervals for μ and σ^2.

First we compute the sample mean

$$\overline{X} = \frac{1}{7}(10.7 + 8.5 + 9.1 + 10.3 + 13.7 + 9.7 + 9.3) = 10.1857$$

and the sample standard deviation

$$S^2 = \frac{1}{6}\left[(10.7 - \overline{X})^2 + \cdots + (9.3 - \overline{X})^2\right] = 2.9448.$$

The general symmetric two-sided $(1-\alpha)100\%$ confidence interval[2] for μ is

$$\overline{X} - t_{\alpha/2}\sqrt{S^2/n} < \mu < \overline{X} - t_{\alpha/2}\sqrt{S^2/n}$$

[1] I didn't explicitly say that the confidence interval needed to be two-sided. It's okay if you gave a one-sided confidence interval $p > \hat{p} - z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ or $p < \hat{p} + z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

[2] Again, it's okay if you used a one-sided confidence interval.

and the only kind of $(1-\alpha)100\%$ confidence interval that we know for σ^2 is

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}.$$

In each case the t and χ^2 values have $n-1$ degrees of freedom. For $\alpha = 5\%$ we have the values

$$t_{\alpha/2}(6) = 2.447, \quad \chi^2_{1-\alpha/2}(6) = 1.237, \quad \chi^2_{\alpha/2}(6) = 14.449.$$

Therefore we obtain the 95% confidence intervals

$$8.60 < \mu < 11.77$$

and

$$1.22 < \sigma^2 < 14.23.$$

Remark: Chi-squared distributions will not be on the exam.