

Probability and Statistics: Counting (permutation and combinations), probability axioms, Sample space, events, independent events, mutually exclusive events, marginal, conditional and joint probability, Bayes Theorem, conditional expectation and variance, mean, median, mode and standard deviation, correlation, and covariance, random variables, discrete random variables and probability mass functions, uniform, Bernoulli, binomial distribution, Continuous random variables and probability distribution function, uniform, exponential, Poisson, normal, standard normal, t-distribution, chi-squared distributions, cumulative distribution function, Conditional PDF, Central limit theorem, confidence interval, z-test, t-test, chi-squared test. — Hypothesis Testing

Wald test → Not in gate syllabus

Statistical Inference

→ Conclusion about parameters of population from the sample set

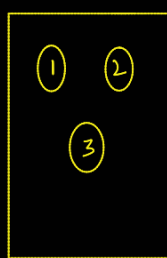
- ① Population → The bigger set
→ Similar to universal set
- ② Sample → subset of population

eg:- Avg salary all employees NVIDIA → parameter of population
But only 200 employee filled form → sample
Based on sample - estimate population avg → statistical inference

- ③ Sample Statistic → Any function of the sample from a population
eg:- Sample mean
Sample SD
Z-score
T-score
 $\bar{X}_{\text{sample}} = f(X_1, X_2, X_3, \dots, X_n)$ $n = \text{elements in one sample}$

- ④ Taking multiple samples to get more accurate result is part of statistical inference
→ Values of sample statistics (eg sample mean) vary for different sample (point estimate)
→ This variation can be defined if we treat the sample statistic as a random variable of the set of possible samples as the domain.

- ⑤ This distribution called is called Sampling distribution of sampling statistics.



$S_1 \rightarrow \{1, 2\} \quad \bar{X}_1 = 1.5$
 $S_2 \rightarrow \{1, 1\} \quad \bar{X}_2 = 1$
 $S_3 \rightarrow \{1, 3\} \quad \bar{X}_3 = 2$

→ Sample statistic → sample mean
Each one individually is called point estimate

*Sampling with Replacement

Sampling distribution → $\bar{X} = f(\text{sample})$
 $P(\bar{X} = x) = P(\text{sample} \in \{s \mid x = f(s)\})$

eg distribution
all 9 point estimates

all 9 samples equally likely

sample	$\bar{X}$
1,1	1
1,2	1.5
1,3	2
2,1	1.5
2,2	2
2,3	2.5
3,1	2
3,2	2.5
3,3	3

$\bar{X}$	$P(\bar{X})$
1	1/9
1.5	2/9
2	3/9
2.5	2/9
3	1/9

Sample distribution PDF

Central Limit Theorem

① If you have a populⁿ with parameters θ , and some distribution with μ, σ
 & let $\bar{x}$ be the point estimate mean, i.e. $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, where $n \rightarrow$ sample size

$$\bar{x} \sim \text{Normal}(\mu_{\bar{x}}, \sigma_{\bar{x}}^2) \text{ as } n \rightarrow \infty$$

② Sample size $\rightarrow$
 • If sample size is small, the variance of sampling distribution will be larger
 • if sample size is large, the variance of sampling distribution will be smaller
 $\rightarrow$ law of large numbers.

eg:- Height of student of 10 classes (X)

• $X \sim K(\mu, \sigma^2)$ k $\rightarrow$ unknown distribution

• Sample mean $\rightarrow \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

• CLT says $\bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$

1 sample = $\{x_1, x_2, x_3, \dots, x_n\}$

1 sample $\Rightarrow$ n size

$n \rightarrow$ sample size
 = # elements in one sample

Find $\mu_{\bar{x}}$ & $\sigma_{\bar{x}}^2$ without CLT

$$\bar{x} = (x_1 + x_2 + \dots + x_n) \cdot \frac{1}{n}$$

$$E(\bar{x}) = \frac{1}{n} E\left(\sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} \cdot n\mu = \mu$$

$$\mu_{\bar{x}} = \mu$$

$$\text{Var}(\bar{x}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i)$$

$$\sigma_{\bar{x}}^2 = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

x_i are independent

Conclusion

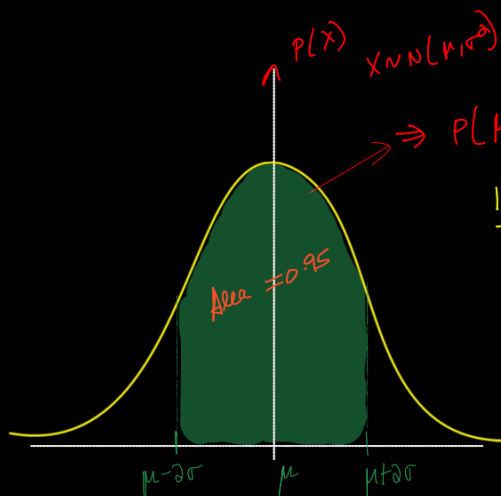
$X \sim K(\mu, \sigma^2)$; $\bar{x} = \frac{1}{n} \sum_{i=1}^n (x_i)$ where $n \rightarrow$ sample size; then

CLT $\Rightarrow \bar{x} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2) \rightarrow$ industry standard if $(n > 30)$

w/o CLT $\Rightarrow \mu_{\bar{x}} = \mu$ & $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \rightarrow$ always

LOLN

$\rightarrow$ for sufficiently large samples $n \rightarrow \infty$ ① we consider $\bar{x} \sim N(\mu, \sigma^2/n)$ has converged to the true value of μ ② since the variance converged to zero



Implication

① 95% of the time our sample mean, i.e. point estimate will lie inside this interval $\mu - 2\sigma < x < \mu + 2\sigma$ if the sample size > 30
 $\rightarrow$ CLT applies for $n > 30$
 $\rightarrow$ This is industry standard.

CLT $\rightarrow$ formal definition

Central Limit Theorem: If $\bar{X}$ is the mean of a random sample of size n taken from a population with mean μ and finite variance σ^2 , then the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

as $n \rightarrow \infty$, is the standard normal distribution $n(z; 0, 1)$.

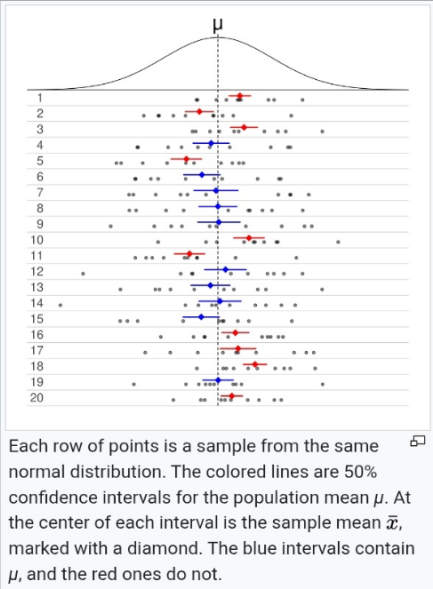
The normal approximation for $\bar{X}$ will generally be good if $n \geq 30$, provided the population distribution is not terribly skewed. If $n < 30$, the approximation is good only if the population is not too different from a normal distribution and, as stated above, if the population is known to be normal, the sampling distribution of $\bar{X}$ will follow a normal distribution exactly, no matter how small the size of the samples

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

$\rightarrow$ How to search such a table
 $\rightarrow$ Row decides most most significant 2 digits which includes 1 decimal place
 $\rightarrow$ Column decides least significant digit $\rightarrow$ 2nd decimal place
 $\rightarrow$ So for eg. $P(Z < 1.23) = 13^{\text{th}}$ row & 4th column
 $P(Z < 1.23) = 0.8907$

CONFIDENCE INTERVAL

A 95% confidence level does not imply a 95% probability that the true parameter lies within a particular calculated interval. The confidence level instead reflects the long-run reliability of the method used to generate the interval.^{[2][3]} In other words, if the same sampling procedure were repeated 100 times from the same population, approximately 95 of the resulting intervals would be expected to contain the true population mean.



Each row of points is a sample from the same normal distribution. The colored lines are 50% confidence intervals for the population mean μ . At the center of each interval is the sample mean $\bar{x}$, marked with a diamond. The blue intervals contain μ , and the red ones do not.

→ Each row (→ 20) represents a sample
 → point estimate $\bar{x}$ → diamond
 → Notice the interval is calculated separately for each sample
 → The length of interval need not be same and depends on the method used to generate this interval

Wrong interpretation
 → Given any interval $[\bar{x}_i - c, \bar{x}_i + c]$ there is a 95% probability for θ to be inside $[\bar{x}_i - c, \bar{x}_i + c]$

→ This is because for each sample the interval is recalculated & even may be of different lengths depending on the method of calculating the interval

Definition 2 Tailed Confidence Interval

Let X be a random sample from a probability distribution with statistical parameter (θ, φ) . Here, θ is the quantity to be estimated, while φ includes other parameters (if any) that determine the distribution. A confidence interval for the parameter θ , with confidence level or coefficient γ , is an interval $(u(X), v(X))$ determined by random variables $u(X)$ and $v(X)$ with the property:

$$P(u(X) < \theta < v(X)) = \gamma \quad \text{for all } (\theta, \varphi).$$

→ Function of Random sample
 → Changes with each sample

the asymptotically standard normal quantity

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

$\bar{X}$ → Sample statistic
 μ → population mean
 S → Sample S.D $\neq \sqrt{\text{Var } \bar{X}}$ → misconception don't confuse

where μ and n are the population mean and the sample size, respectively.

$$\sqrt{\text{Var}(\bar{X})} = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} + s \quad \left| \begin{array}{l} \sigma = \text{population S.D} \\ s = \text{sample S.D} \end{array} \right.$$

if sample k has n points $(x_1, x_2, x_3, \dots, x_n)$ then

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n (x_i)$$

$$s_k^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_k)^2$$

Suppose $X_1, \dots, X_n$ is an independent sample from a normally distributed population with unknown parameters mean μ and variance σ^2 . Define the sample mean $\bar{X}$ and unbiased sample variance S^2 as

$$\bar{X} = \frac{1}{n} (X_1 + \dots + X_n),$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Then the value

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has a Student's t distribution with $n - 1$ degrees of freedom.^[5] This value is useful because its distribution does not depend on the values of the unobservable parameters μ and σ^2 ; i.e., it is a pivotal quantity.

→ large interval → high variance sample
 → small interval → low variance sample

Suppose we wanted to calculate a 95% confidence interval for μ . First, let c be the 97.5th percentile of the distribution of T . Then there is a 2.5% chance that T will be less than $-c$ and a 2.5% chance that it will be larger than $+c$ (as the t distribution is symmetric about 0). In other words,

$$P_T(-c \leq T \leq c) = 0.95. \quad \rightarrow \text{To find } c \rightarrow P(T \leq c) = 1 - \alpha/2 = 0.975 \rightarrow \text{look up Table}$$

Consequently, by replacing T with $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ and re-arranging terms,

$$P_X \left(\bar{X} - \frac{cS}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{cS}{\sqrt{n}} \right) = 0.95$$

where P_X is the probability measure for the sample $X_1, \dots, X_n$.

$$P(u(\bar{x}) \leq \theta \leq v(\bar{x})) = \gamma$$

It means that there is 95% probability with which this condition $\bar{X} - \frac{cS}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{cS}{\sqrt{n}}$ occurs in repeated sampling. After observing a sample, we find values $\bar{x}$ for $\bar{X}$ and s for S , from which we compute the below interval, and we say it is a 95% confidence interval for the mean.

$$\left[\bar{x} - \frac{cs}{\sqrt{n}}, \bar{x} + \frac{cs}{\sqrt{n}} \right].$$

Interpretation [edit]

Various interpretations of a confidence interval can be given (taking the 95% confidence interval as an example in the following).

- The confidence interval can be expressed in terms of a **long-run frequency** in **repeated samples** (or in **resampling**): "Were this procedure^[1] to be repeated on numerous samples, the proportion of calculated 95% confidence intervals that encompassed the true value of the population parameter would tend toward 95%."^[6]
- The confidence interval can be expressed in terms of probability with respect to a single theoretical (yet to be realized) sample: "There is a 95% **probability** that the 95% confidence interval calculated from a given future sample will cover the true value of the population parameter."^[7] This essentially reframes the "repeated samples" interpretation as a probability rather than a frequency.
- The confidence interval can be expressed in terms of statistical significance, e.g.: "The 95% confidence interval represents values that are not **statistically significantly** different from the point estimate at the .05 level."^[8]

2 → This is however not true for any particular sample that is already observed & for which the interval is calculated.

2 → This is because $P(u(x) \leq \theta \leq v(x)) = \gamma$ is only true for x unobserved → The moment x is observed (sampling done) the $u(x)$ & $v(x)$ just becomes calculated margins.

2 → This means the statement 2 is false when treating x as a RV that is yet to be realised or observed in future.

$P(u(x) \leq \theta \leq v(x) | x) \neq \gamma \rightarrow \star$

1) The procedure is what repeats across samples, not the interval & not even the interval length

→ This procedure gives different intervals of different sizes for each sample that we repeat

2) When talking about the future sample and its calculated confidence interval it is true that θ as a population parameter will lie inside this future interval with probability 95%

Common misunderstandings [edit]

Confidence intervals and levels are frequently misunderstood, and published studies have shown that even professional scientists often misinterpret them.^[9]

- A 95% confidence level does not mean that for a given realized interval there is a 95% probability that the population parameter lies within the interval.^{[10][11]}
- A 95% confidence level does not mean that 95% of the sample data lie within the confidence interval.^[1]
- A 95% confidence level does not mean that there is a 95% probability of the parameter estimate from a repeat of the experiment falling within the confidence interval computed from a given experiment.^[11]

For example, suppose a factory produces metal rods. A random sample of 25 rods gives a 95% confidence interval for the population mean length of 36.8 to 39.0 mm.^[12]

- It is incorrect to say that there is a 95% probability that the true population mean lies within this interval, because the true mean is fixed, not random. For example, it might be 37 mm, which is within the confidence interval, or 40 mm, which is not; in any case, whether it falls between 36.8 and 39.0 mm is a matter of fact, not probability.
- It is not necessarily true that the lengths of 95% of the sampled rods lie within this interval. In this case, it cannot be true: 95% of 25 is not an integer.
- It is incorrect to say that if we took a second sample, there is a 95% probability that the sample mean length (an estimate of the population mean length) would fall within this interval. In fact, if the true mean length is far from this specific confidence interval, it could be very unlikely that the next sample mean falls within the interval.

Instead, the 95% confidence level means that if we took 100 such samples, we would expect the true population mean to lie within approximately 95 of the calculated intervals.^{[1][10][11][12]}

Question solving

→ $X = \text{sample}$

$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$

$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

LCLT as $n > 30$

1 method

$P(-c \leq Z \leq +c) \geq 1 - \alpha$

$P(Z \leq c) = 1 - \alpha + \alpha/2$

Find c from Z table

$P(-c \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq +c)$

Being back in terms of X

$P(-c \leq Z \leq c) = P(\bar{X} - \frac{c\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{c\sigma}{\sqrt{n}}) = P(\mu - \frac{c\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + \frac{c\sigma}{\sqrt{n}}) \rightarrow \text{when } X \text{ unobserved}$

→ The moment we plug in the value of $\bar{X}$ as an observation X then the probability interpretation fails. Since there is no random variable inside the probability function. Even though μ is unknown. It is not a RV. just a constant.

→ Wrong interpretation comes when μ is treated like RV.

→ Think that your friends know μ , but you only don't know

where c is critical value obtained from the table for this statistic.

↙ In question they might use notation z^*

Standard error (SE)

for any statistic $\bar{H}$

$[\bar{H} - c(SE), \bar{H} + c(SE)]$

is how to calculate confidence interval

The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter.

99% confidence interval for $\bar{X}_1 = \left[\bar{X}_1 - 2.57 \cdot \frac{0.3}{\sqrt{36}} \leq \mu \leq \bar{X}_1 + 2.57 \cdot \frac{0.3}{\sqrt{36}} \right]$

$\bar{X}_1 = \left[2.6 - 0.1285 \leq \mu \leq 2.6 + 0.1285 \right]$

$\bar{X}_1 = \left[2.4715 \leq \mu \leq 2.7285 \right]$ 99%

$n = 36$
 $\bar{X}_1 = 2.6 \text{ g/mL}$
 given $\sigma = 0.3 \text{ g/mL}$
 $\sigma = \text{population SD}$
 $\bar{X} = \text{sample mean (statistic)}$
 95% confidence interval = $\left[\bar{X}_1 - 1.96 \times \frac{0.3}{\sqrt{36}} \leq \mu \leq \bar{X}_1 + 1.96 \times \frac{0.3}{\sqrt{36}} \right]$
 = $\left[2.6 - 0.098 \leq \mu \leq 2.6 + 0.098 \right]$
 = $\left[2.502 \leq \mu \leq 2.698 \right]$

One Tailed Confidence Interval

Example 9.4: In a psychological testing experiment, 25 subjects are selected randomly and their reaction time, in seconds, to a particular stimulus is measured. Past experience suggests that the variance in reaction times to these types of stimuli is 4 sec² and that the distribution of reaction times is approximately normal. The average time for the subjects is 6.2 seconds. Give an upper 95% bound for the mean reaction time.

Solution: The upper 95% bound is given by

$$\bar{x} + z_{\alpha} \sigma / \sqrt{n} = 6.2 + (1.645) \sqrt{4/25} = 6.2 + 0.658 = 6.858 \text{ seconds.}$$

Hence, we are 95% confident that the mean reaction time is less than 6.858 seconds.

population variance = 4
 SD = 2
 $\Rightarrow$ w/o CLT also sample mean $\sim N(\mu_x, \sigma_x^2)$
 $\bar{X}_1$
 $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$
 $\star \Rightarrow P\left(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} > -c\right) = 1 - \alpha = 0.95$ | $c = -1.645$
 $c > 0$
 95% upper confidence interval = $P\left[\mu \leq \bar{X} + \frac{c\sigma}{\sqrt{n}}\right] = 0.95$
 $P\left(\mu \leq 6.2 + \frac{(1.645 \times 2)}{5}\right) = 0.95$

8.67 The breaking strength X of a certain rivet used in a machine engine has a mean 5000 psi and standard deviation 400 psi. A random sample of 36 rivets is taken. Consider the distribution of $\bar{X}$, the sample mean breaking strength.

(a) What is the probability that the sample mean falls between 4800 psi and 5200 psi?

(b) What sample n would be necessary in order to have

$$P(4900 < \bar{X} < 5100) = 0.99?$$

b) c for $P(-c \leq Z \leq c) = 0.99$

$c = 2.57$

We need $100 = c \frac{\sigma}{\sqrt{n}}$

$\Rightarrow n \geq \left(2.57 \times \frac{400}{100}\right)^2 = (2.57 \times 4)^2$

$n \geq 105.67$

$n = 106$

Population
 $= n$

a) $P\left(\mu - c \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + c \cdot \frac{\sigma}{\sqrt{n}}\right) = P(-c \leq Z \leq c)$

$\mu = 5000$

$c \frac{\sigma}{\sqrt{n}} = 200 \Rightarrow c = \frac{200 \times 6}{400} = 3$

Find $P(Z \leq c) = k$ i.e. $P(Z \leq 3)$

then our confidence for given interval = $k - (-k) = 2k - 1$

$k = 0.9987$

confidence = 99.74%

γ -Confidence Interval = Statistic $\pm$ (critical value * standard error) $\rightarrow \left[\bar{X} \pm z^* \left(\frac{\sigma}{\sqrt{n}} \right) \right]$

using Z-score statistic

① $\bar{X} \pm \left(\frac{\sigma}{\sqrt{n}} \right) z^*$

$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

$P(-z^* \leq Z \leq z^*) = \gamma$

Using CLT $n > 30$

w/o CLT if sampling from $N(\mu, \sigma^2)$

$\rightarrow$ Can use $S \rightarrow$ sample SD as proxy for σ is $n > 30$

using T-score statistic

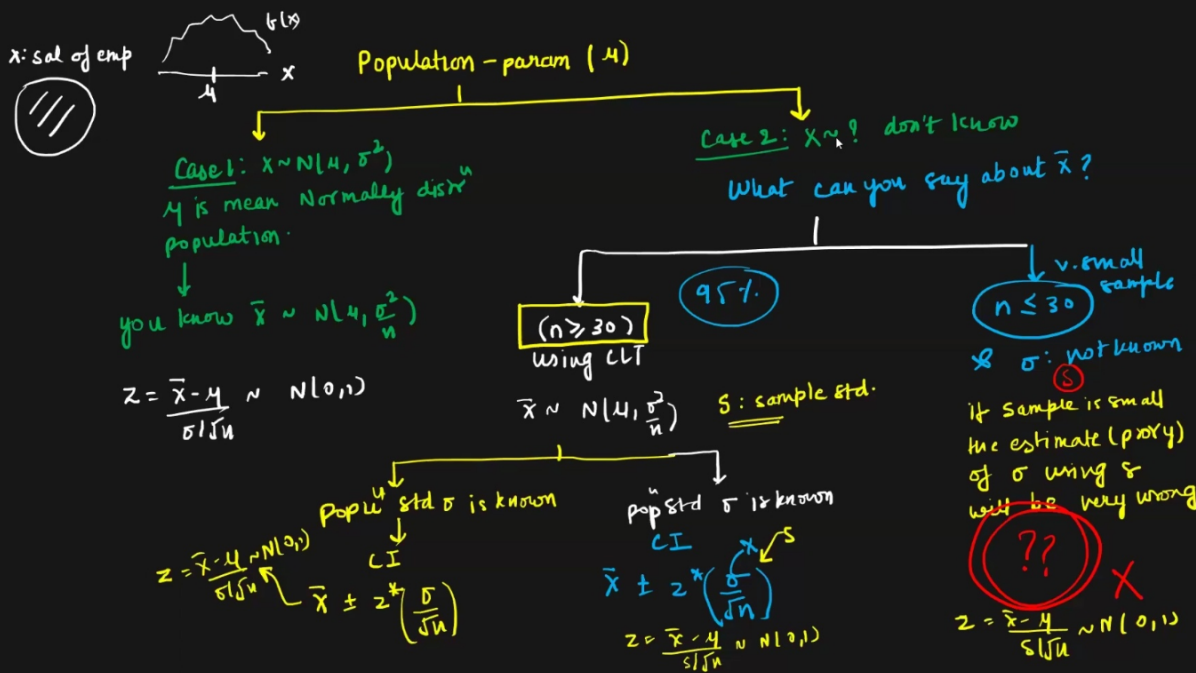
② $\bar{X} \pm \left(\frac{S}{\sqrt{n}} \right) t^*$

$T = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim T(n-1)$

$P(-t^* \leq T \leq t^*) = \gamma$

$\rightarrow n \leq 30 \rightarrow$ we can't use S as proxy for σ
 $\rightarrow$ In that case if sampling from $N(\text{unknown } \mu, \sigma^2)$

Student - t distribution



Assume, popul^y of interest has normal distribution with unknown mean μ & unknown variance σ^2 . Assume you take $x_1, \dots, x_n$ (n -samples)
 let, $\bar{X}$ & s^2 be sample mean & sample variance.

↓
 samp std = s

We want to construct 2 side CI?

(n is not sufficient) or σ is not known.

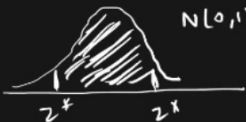
↓
 σ^2 -known

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

↓
 σ^2 -unknown

$$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \neq N(0,1)$$

$Z \sim t$ -distribution with $(n-1)$ degree of freedom



CI: $\bar{X} \pm z^* SE$

Probability density function

Student's t-distribution has the probability density function

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

where ν is the number of degrees of freedom and Γ is the Gamma function.

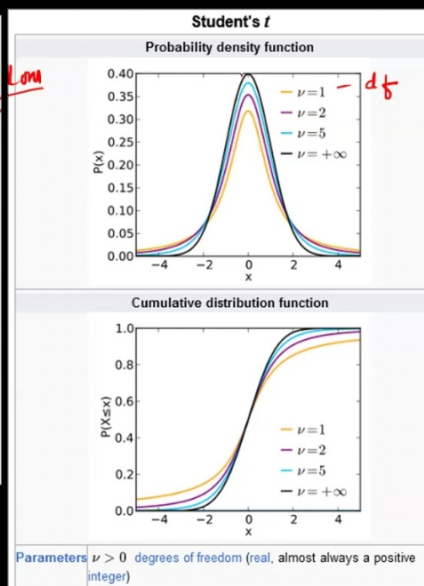
For ν even,

$$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} = \frac{(\nu-1)(\nu-3)\dots 5 \cdot 3}{2\sqrt{\nu}(\nu-2)(\nu-4)\dots 4 \cdot 2}$$

For ν odd,

$$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} = \frac{(\nu-1)(\nu-3)\dots 4 \cdot 2}{\pi\sqrt{\nu}(\nu-2)(\nu-4)\dots 5 \cdot 3}$$

not remember
 ν : degree of freedom



Normal, chi-squared, gamma, T-distribution

$$V \sim \chi_n^2 = \text{gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

$$\chi_n^2 = Z_1^2 + Z_2^2 + Z_3^2 + \dots + Z_n^2 \quad \text{where } Z_1, \dots, Z_n \sim N(0,1) \text{ iid RVs}$$

$$\chi_1^2 = Z^2 = \text{gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$$

$$E(\chi_n^2) = n$$

$$\text{Var}(\chi_1^2) = \text{Var}(Z^2) = \text{Var}(\text{gamma}\left(\frac{1}{2}, \frac{1}{2}\right)) = 2 //$$

$$\text{Var}(\chi_n^2) = \text{Var}(\text{gamma}\left(\frac{n}{2}, \frac{1}{2}\right)) = \frac{n/2}{(1/2)^2} = \underline{\underline{2n}}$$

$$= \text{Var}(Z_1^2 + Z_2^2 + \dots + Z_n^2) \\ = n \text{Var}(Z^2) = \underline{\underline{2n}} \quad E(V_n^k) = \prod_{j=0}^{k-1} (n+2j)$$

$$T_n = \frac{Z}{\sqrt{V/n}} \quad \begin{array}{l} Z \sim N(0,1) \\ V \sim \chi_n^2 \end{array} \quad \left| \begin{array}{l} V \text{ \& } Z \text{ independent} \\ n \rightarrow \text{degree of freedom} \end{array} \right.$$

→ Symmetric about 0

$$\rightarrow T_1 = \frac{Z_1}{|Z_2|} \sim \text{Cauchy's distribution} \quad Z_1, Z_2 \text{ iid } N(0,1)$$

$$E(T_1) = \text{infinite} \rightarrow \text{not defined}$$

$$\rightarrow n \geq 2 \Rightarrow E(T_n) = E(Z) \cdot E\left(\sqrt{\frac{n}{V}}\right) \quad \begin{array}{l} \rightarrow \text{independence} \\ \Rightarrow \text{uncorrelated} \end{array}$$

$$E(Z^2) = 1 = \text{Var}(Z) = E(Z^2) - E^2(Z)$$

$$E(Z^4) = 3 \quad \left. \begin{array}{l} E(Z^3) = 0 \\ E(Z^5) = 0 \end{array} \right\} \text{odd powers} = 0$$

$$E(Z^6) = 15$$

$$E(Z^{2n}) = \text{skip factorial} = 1 \times 3 \times 5 \times \dots \times (2n-1)$$

$$E[Z^{2n}] = E[(Z^2)^n] \quad \text{let } Z^2 = h \\ h \sim \text{gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

$$= E(h^n) = \frac{\frac{1}{2} \cdot \frac{1}{2} + 1}{\frac{1}{2}} \cdot \frac{\frac{1}{2} + 1}{\frac{1}{2}} \cdot \frac{\frac{1}{2} + 2}{\frac{1}{2}} \dots \frac{\frac{1}{2} + (n-1)}{\frac{1}{2}}$$

$$E(Z^{2n}) = E(h^n) = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-1)$$

$$E(Z^k) = \begin{cases} 1 \cdot 3 \cdot 5 \cdot \dots \cdot (k-1) & k \text{ even} \\ 0 & k \text{ odd} \end{cases}$$

$$h \sim \text{gamma}(a, 1) = \frac{1}{\Gamma(a)} (dx)^a \frac{e^{-dx}}{x}$$

$$\Gamma(a) = \int_0^\infty x^a \frac{e^{-x}}{x} dx$$

$$\Gamma(a+1) = a \Gamma(a)$$

$$\Gamma(n) = (n-1)! \quad n = \text{pos. Integer}$$

$$\Gamma(1/2) = \sqrt{\pi}$$

$$\Gamma(3/2) = \frac{\sqrt{\pi}}{2}$$

$$\Gamma(5/2) = \frac{3}{4} \sqrt{\pi}$$

$$E(h) = \frac{a}{1} \quad \text{Var}(h) = \frac{a}{1^2}$$

$$\text{In general } E(h^k) = \frac{\Gamma(a+k)}{\Gamma(a)} = \frac{a(a+1)}{1 \cdot 1} \frac{(a+2)}{1} \frac{(a+k-1)}{1}$$

variance (T_n)

$$= E(T_n^2) - E^2(T_n) \quad \left\{ \begin{array}{l} \text{not defined, } n=1 \\ =0, \quad n>1 \end{array} \right.$$

$$E(T_n^2) = \frac{E(Z^2)}{1} \cdot E\left(\frac{n}{V}\right)$$

$$E(T_n^2) = n E\left(\frac{1}{V}\right)$$

$$V \sim \chi_n^2$$

$$V \sim \text{gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

$$E(T_n^2) = n \cdot \frac{1}{\Gamma(n/2)} \int \frac{1}{v} \cdot (1/2)^{n/2} \cdot e^{-v/2} \frac{dv}{v}$$

$$E(T_n^2) = \frac{n}{\Gamma(n/2)} \cdot \frac{1}{2} \int \left(\frac{v}{2}\right)^{n/2-1} \cdot e^{-v/2} \frac{dv}{v} \quad \begin{array}{l} \text{put } \frac{v}{2} = x \\ dv = 2dx \end{array}$$

$$E(T_n^2) = \frac{n}{2 \Gamma(n/2)} \int x^{(n/2)-1} \frac{e^{-x}}{x} dx = \frac{n}{2} \cdot \frac{\Gamma(n/2-1)}{\Gamma(n/2)}$$

$$E(T_n^2) = \frac{n}{2} \cdot \frac{\Gamma(n/2-1)}{(n/2-1) \Gamma(n/2-1)} = \frac{n}{2} \cdot \frac{2}{n-2}$$

$$\Rightarrow E(T_n^2) = \frac{n}{n-2}$$

$$\Rightarrow \text{Var}(T_n) = \begin{cases} \frac{n}{n-2} & n > 2 \\ \text{undefined} & n = 1, 2 \end{cases}$$

n^{th} moment of T_n doesn't exist
odd moment of T_n if exist $= 0$

Let $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$ if $X_j \sim \text{Normal}$ Then $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ degree of freedom only if $X_j \sim \text{Normal}$

$S^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$ $\bar{X}$ — Sample mean $S \rightarrow$ sample SD
 $S \rightarrow$ approximation of σ as $n \rightarrow \infty$
 $S/\sqrt{n} \rightarrow$ SD of $\bar{X}$ if $\bar{X}$ is an RV

1) degree of freedom intuition $\rightarrow$ S is having $n-1$ degree of freedom since not all X_j are free to vary given sample mean $\bar{X}$. So if we calculate variance by normal equation $\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ then we are underestimating actual variance

2) To understand think of the extreme case when $n=2$, if we sample size is only 2, then knowing the mean & one of the value is sufficient to determine the other value. So degree of freedom given mean & n sample size is actually $n-1$

cum. prob	$t_{.50}$	$t_{.25}$	$t_{.20}$	$t_{.15}$	$t_{.10}$	$t_{.05}$	$t_{.025}$	$t_{.01}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
Z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Reading table

Ignacio is curious about the average age of cars in the commuter lot at his large university. He takes a random sample of 16 cars and finds that their average age is $\bar{x} = 14.5$ years and their standard deviation is $s_x = 4.6$ years. The distribution of ages in the sample was roughly symmetric with no obvious outliers.

Based on this sample, which of the following is a 90% confidence interval for the mean age of cars (in years) in this commuter lot?

Choose 1 answer:

- (A) (6.44, 22.56)
- (B) (12.05, 16.95)
- ☒ (C) (12.48, 16.52)
- (D) (12.61, 16.39)
- (E) (13.35, 15.65)

$n=16$
 $\bar{x}_1 = 14.5$
 $S_1 = 4.6$

3 20
 1753 x
 1.15
 8765
 1753 x
 1753 x x
 201595 2.00

degree of freedom = 15 $\Rightarrow t^* = 1.753$

$\left[\bar{x}_1 - t^* \frac{S_1}{\sqrt{n}} \leq \mu \leq \bar{x}_2 + t^* \frac{S_1}{\sqrt{n}} \right]_{90\%}$ for sample 1

$\left[14.5 - \frac{1.753 \times 4.6}{4} \leq \mu \leq 14.5 + \frac{1.753 \times 4.6}{4} \right]_{90\%}$ for sample 1
 $= [12.48 \leq \mu \leq 16.52]_{90\%}$ for sample 1

7 patients' blood pressures have been measured after having been given a new drug for 3 months. They had blood pressure increases of 1.5, 2.9, 0.9, 3.9, 3.2, 2.1 and 1.9. Construct a 95% confidence interval for the true expected blood pressure increase for all patients in a population.

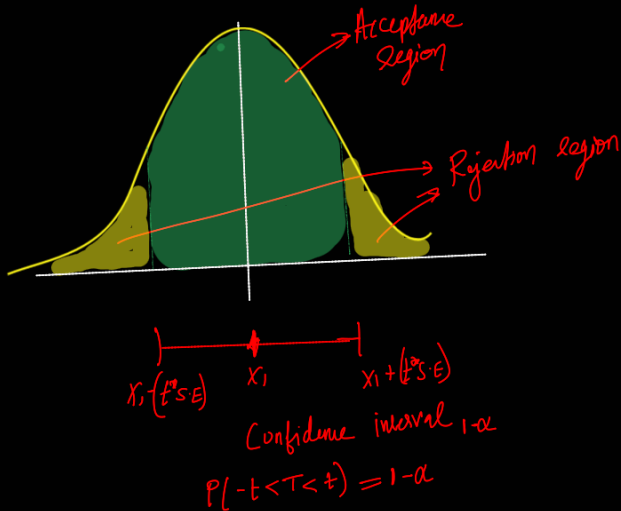
$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_7$
 (since $t < 30$)

$t_{\alpha/2, n-1} = 2.447$
 $DOF = 6$

$\left[\bar{X} - 2.447 \frac{S_1}{\sqrt{n}} \leq \mu \leq \bar{X} + 2.447 \frac{S_1}{\sqrt{n}} \right]_{90\%}$ for sample 1

$\bar{x}_1 = \frac{1.5 + 2.9 + 0.9 + 3.9 + 3.2 + 2.1 + 1.9}{7}$
 $S_1 = \frac{1}{6} \sum_{i=1}^7 (x_i - \bar{x})^2$

Hypothesis testing



Statistical hypothesis $\rightarrow$ Statement about parameter of one or more population
 $\rightarrow$ Not statements about sample $\star$
 eg:- Avg salary of populⁿ = 20L $\rightarrow H_0: \mu = 20L$ (Null hypothesis)
 Avg salary of populⁿ $\neq$ 20L $\rightarrow H_1: \mu \neq 20L$ (Alternate hypo)

Null hypothesis (H_0)
 $\rightarrow$ Based on prev data or domain expert

Alternate hypothesis (H_a)
 $\rightarrow$ Opposite to Null hypothesis
 $\rightarrow$ Can be two tailed or one tailed.

- Parameter of interest:** From the problem context, identify the parameter of interest.
- Null hypothesis, H_0 :** State the null hypothesis, H_0 .
- Alternative hypothesis, H_a :** Specify an appropriate alternative hypothesis, H_1 .
- Test statistic:** Determine an appropriate test statistic. 3
- Reject H_0 if:** State the rejection criteria for the null hypothesis. 3
- Computations:** Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.
- Draw conclusions:** Decide whether or not H_0 should be rejected and report that in the problem context.

Z-score $\rightarrow Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \rightarrow$ (n large or sampling from normal)
 T-score $\rightarrow T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \rightarrow$ (n small & sampling from normal)

$\rightarrow$ ① Significance level (α)
 $\rightarrow$ ② P-value

p-value = P(A more extreme than observed / Null hypothesis is true)

if $p\text{-value} < \alpha \rightarrow$ reject H_0
 $p\text{-value} \geq \alpha \rightarrow$ fail to reject H_0

$\alpha \rightarrow$ Significance level

$\rightarrow$ There is no accepting in hypothesis testing
 $\rightarrow$ There is only $\rightarrow$ fail to reject $\Rightarrow$ Not enough extreme data observed to reject

A certain bag of fertilizer advertises that it contains 7.25 kg, but the amounts these bags actually contain is normally distributed with a mean of 7.4 kg and a standard deviation of 0.15 kg.

The company installed new filling machines, and they wanted to perform a test to see if the mean amount in these bags had changed. Their hypotheses were

$H_0: \mu = 7.4 \text{ kg}$
 $H_a: \mu \neq 7.4 \text{ kg}$

They took a random sample of 50 bags and observed a sample mean and standard deviation of $\bar{x} = 7.36 \text{ kg}$ and $s_x = 0.12 \text{ kg}$. They calculated that these results had a P-value of approximately 0.02.

What conclusion should be made using a significance level of $\alpha = 0.05$:

Choose 1 answer:

- (A) Fail to reject H_0
- (B) Reject H_0 and accept H_a
- (C) Accept H_0

$p < \alpha \Rightarrow$ reject H_0 accept H_a

P-value in hypothesis testing is defined as the probability for the observed sample given the null hypothesis is true. But why do we use this probability instead of using the more natural probability of hypothesis being true given the observed sample?



This is a very good question, and it gets to the philosophical core of hypothesis testing.

Let me first restate the contrast clearly:

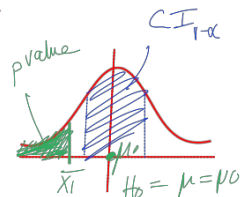
- What a p-value is:

$$p\text{-value} \neq P(\text{data as extreme as observed} \mid H_0 \text{ is true})$$

- What feels more natural:

$$P(H_0 \text{ is true} \mid \text{observed data})$$

$P(\text{observed data} \mid H_0) \rightarrow$ Since continuous this will be zero
 $\rightarrow$ We need area under curve
 $\rightarrow$ gives area under curve for all values as more extreme than observed



p value =
 smallest α that still rejects H_0
 largest α that that fails to reject H_0

These are **not the same**, and the reason we use the first instead of the second is a mix of **logic, practicality, and historical philosophy**.

1. Why the “natural” probability is not directly available

To compute

$$P(H_0 \mid \text{data}),$$

you *must* use **Bayes’ theorem**:

$$P(H_0 \mid D) = \frac{P(D \mid H_0) P(H_0)}{P(D)}$$

This requires:

1. A **prior probability** $P(H_0)$
2. A probability model for *all* competing hypotheses

In classical (frequentist) statistics:

- H_0 is **not a random variable**
- It is either true or false
- Therefore $P(H_0)$ is *not defined*

So frequentist inference **cannot** talk about
“probability that the hypothesis is true.”

👉 This is not a technical limitation—it’s a *philosophical choice*.

2. Frequentist philosophy: probabilities belong to data, not hypotheses

Frequentist statistics adopts this principle:

Probability describes long-run behavior of random data, not uncertainty about fixed hypotheses.

- Data are random
- Hypotheses are fixed statements about the world

So it makes sense to ask:

$$P(\text{data} \mid H_0)$$

but **not**:

$$P(H_0 \mid \text{data})$$

This is why p-values are framed as:

“If the null were true, how surprising would this data be?”

3. Why this is still useful (despite not being intuitive)

The p-value answers a **decision-relevant question**:

Is the observed data inconsistent with what we would expect if H_0 were true?

If:

- p-value is small → data is **very unlikely** under H_0
- Therefore → treating H_0 as true leads to implausible outcomes

This supports **rejection**, not belief.

Important:

✗ p-value does **not** say “ H_0 is probably false”

✓ it says “this data would be rare if H_0 were true”

4. Why we don't just switch to $P(H_0 \mid \text{data})$

To do that, we must:

- Assign a prior probability to hypotheses
- Model uncertainty over parameter values
- Accept subjectivity in priors

This leads to **Bayesian statistics**, where:

$$P(H_0 \mid \text{data})$$

is exactly what is computed

So your “more natural probability” *does exist* — it’s just **Bayesian**, not frequentist.

5. Why p-values became dominant historically

Three main reasons:

(a) No priors required

- Early statistics (Fisher, Neyman–Pearson) wanted **objective procedures**
- Priors were seen as subjective and controversial

(b) Error control guarantees

Frequentist testing gives long-run guarantees:

- Type I error $\leq \alpha$
- Reproducibility across repeated experiments

Bayesian posterior probabilities **do not** give these guarantees unless carefully calibrated.

(c) Simplicity and universality

- p-values can be computed from data alone
 - No need to agree on prior beliefs
-

6. Key intuition to remember

A useful mental model:

p-value $\neq$ probability the null is true

p-value = measure of conflict between data and null

Think of a courtroom analogy:

- Null hypothesis: *"The defendant is innocent"*
- p-value: *"How surprising is this evidence if the defendant were innocent?"*

We do **not** compute:

"Probability the defendant is innocent given the evidence"

unless we introduce priors (Bayesian reasoning).

7. Summary (one-line answers)

- We use $P(\text{data} \mid H_0)$ because **frequentist statistics forbids probabilities over hypotheses**
- $P(H_0 \mid \text{data})$ requires **Bayesian inference and priors**
- p-values measure **evidence against a hypothesis**, not belief in it
- Your intuition is correct — it just belongs to a **different statistical framework**

If you want, I can next:

- Show a **numerical example** where p-value and posterior probability strongly disagree
- Explain **why p-values are often misinterpreted**
- Compare **p-values vs Bayes factors** (very relevant to this question)

Just tell me.



Hypothesis testing is usually taught in colleges just after the topic confidence interval. Is there any relation for hypothesis testing p-value or hypothesis testing significance alpha with confidence intervals?



Yes — there is a **deep, exact, and structural connection** between **confidence intervals (CIs)** and **hypothesis tests**. They are *two views of the same inferential machinery*. The fact that they are taught back-to-back is not accidental.

I'll explain this in increasing depth.

1. The core equivalence (most important fact)

For standard frequentist procedures:

A two-sided hypothesis test at significance level α rejects H_0

$\Leftrightarrow$

the $(1 - \alpha)$ confidence interval does *not* contain the null value.

Example

- $H_0 : \mu = \mu_0$
- Significance level: $\alpha = 0.05$
- Construct a 95% confidence interval for μ

Then:

- If $\mu_0 \notin \text{CI} \rightarrow \text{reject } H_0$
- If $\mu_0 \in \text{CI} \rightarrow \text{fail to reject } H_0 \neq \text{accept } H_0 \rightarrow \text{wrong decision rule}$

So:

$\boxed{\text{95\% CI} \rightarrow \text{5\% hypothesis test}}$

2. How the p-value fits into this

The **p-value** answers:

What is the smallest α at which H_0 would be rejected?

Now connect this to confidence intervals:

The p-value is the smallest α such that the $(1 - \alpha)$ confidence interval excludes the null value.

Equivalently:

- If you keep shrinking the CI (lowering confidence level)
- The moment the null value drops out
- That confidence level corresponds to the p-value

This is a *very precise mathematical relationship*.

3. Why this equivalence holds (intuition)

Both CIs and hypothesis tests are built from the **same pivot quantity**:

$$\text{Test statistic} = \frac{\text{estimate} - \text{hypothesized value}}{\text{standard error}}$$

Hypothesis test

Reject H_0 if:

$$|\text{statistic}| > z_{\alpha/2}$$

Confidence interval

Include all values satisfying:

$$|\text{statistic}| \leq z_{\alpha/2}$$

So:

- Hypothesis testing asks
“Is the null value outside the plausible range?”
- Confidence intervals give you
“The entire plausible range.”

Same inequality, viewed from opposite directions.

4. One-sided tests vs confidence intervals

The correspondence still holds, but you must match them correctly.

Hypothesis test	Matching CI
Two-sided test at α	Two-sided $(1 - \alpha)$ CI
Upper-tailed test at α	Lower $(1 - \alpha)$ one-sided CI
Lower-tailed test at α	Upper $(1 - \alpha)$ one-sided CI

 Many confusions arise from mismatching these.

5. Significance level α as “CI coverage”

Interpretation link:

- α = probability of rejecting a **true** H_0
- $1 - \alpha$ = probability the CI contains the **true** parameter

These are the *same error control*, expressed differently:

Concept	Error controlled
Hypothesis testing	Type I error
Confidence interval	Coverage probability

6. Why confidence intervals are often better than p-values

CIs give **strictly more information**:

- Hypothesis test: *reject / not reject*
- p-value: *degree of incompatibility*
- CI: *range of plausible parameter values*

Example:

- p-value = 0.04 → “significant”
- CI: [0.01, 0.02] vs [0.0001, 10]
→ radically different scientific meaning

Same p-value, wildly different implications.

This is why modern statistics often recommends:

Report confidence intervals instead of (or along with) p-values.

7. Important caveat: this equivalence is not universal

The clean equivalence holds when:

- Tests are based on **inversion of CIs**
- Regular models (normal, t, large-sample)
- No discrete complications

It can fail or become approximate for:

- Discrete distributions (binomial, Poisson)
- Composite nulls
- Non-standard or exact tests

But conceptually, the link remains.

8. Big picture summary

- Confidence intervals and hypothesis tests are **dual** concepts
 - A $(1 - \alpha)$ CI contains exactly the values **not rejected** by an α test
 - p-value tells you **how much confidence you must give up** to exclude the null
 - This is why they are taught together — they are two sides of the same coin
-

If you want, I can:

- Show a **geometric picture** of CI vs test
- Walk through a **full numerical example**
- Explain **why “fail to reject” $\neq$ “accept”** using CIs
- Connect this to **likelihood and Bayesian credible intervals**

Just tell me which one you'd like.



Confidence interval has no usefull meaning for a already observed sample statistics. Confidence interval has meaning for the long running sample statistics. How can we then formulate the confidence interval to reject or accept the null hypothesis just based on one sample. It needs to be the long running sample statistics that accept or reject a hypothesis



This is an excellent objection — and you are *thinking exactly like a careful frequentist*. What you are noticing is **the central conceptual tension** in confidence intervals and hypothesis tests.

Short answer:



We do not interpret a confidence interval as a probability statement about the observed sample.

We use it as a decision rule whose *long-run behavior* is guaranteed.

Let me unpack this carefully.

1. You are absolutely right about confidence intervals

In frequentist statistics:

- The **parameter** θ is fixed (unknown)
- The **confidence interval** is random (because it depends on the sample)

Formally:

$$P_{\theta}(\theta \in CI(X)) = 1 - \alpha$$

This probability is about **repeated sampling**, *not* about the specific interval you observed.

So statements like:

“There is a 95% probability that θ lies in this interval”

are **incorrect** in frequentist logic.

✓ Your understanding here is 100% correct.

2. Then how can one single CI justify rejecting a hypothesis?

This is the key point:



We do not use the CI as a statement of belief about this sample.
We use it as a realization of a procedure with guaranteed long-run error control.

Analogy (very important)

Think of a **court rule**, not a belief statement.

- The legal rule might be:
“Convict if evidence exceeds threshold T”
- That rule is justified because:

“In the long run, this rule wrongly convicts $\leq 5\%$ of innocent people”

For *this one defendant*:

- They are either innocent or not
- The probability statement does **not** apply to them individually

→ Hypothesis testing
may not be true
when applied indi-
vidually to a
sample, but in the
long run it is true
(1- α) % of times

Yet the decision is still made using that rule.

👉 Confidence-interval-based testing works the same way.

3. CI inversion is a decision rule, not a probability claim

When we say:

“Reject $H_0 : \theta = \theta_0$ if $\theta_0 \notin CI$ ”

we are saying:

“Use a procedure whose **long-run Type I error rate** is α .”

The CI is just a **computational device** that encodes this rule.

4. Where the long-run guarantee comes from

Let $CI(X)$ be a $(1 - \alpha)$ confidence interval.

Then:

$$P_{\theta_0}(\theta_0 \notin CI(X)) = \alpha$$

This is **exactly** the Type I error definition:

$$P_{\theta_0}(\text{reject } H_0) = \alpha$$

So the rejection rule:

$$\text{Reject if } \theta_0 \notin CI(X)$$

is **frequentist-valid**, even though it uses one observed sample.

5. Why “accept / reject” language is misleading

You are also implicitly pointing out another subtlety:

- Frequentist tests do **not** “accept” hypotheses
- They only **reject** or **fail to reject**

Why?

Because:

- A wide CI containing θ_0 does **not** provide evidence for H_0
- It only means the data is not inconsistent enough

Confidence intervals make this clearer than p-values:

- If CI is wide → uncertainty is high
 - “Fail to reject” is a statement about lack of precision, not truth
-

6. Why one sample is enough (conceptually)

The key idea:

Frequentist guarantees are about the procedure, not the outcome.

- You design a rule using probability theory
- You apply it once
- The guarantee holds even if you observe only one realization

→ Each decision may not be perfect.
→ But when the same procedure of rejecting H_0 is applied, the long run statistics will be guaranteed to be correct in $(1-\alpha)\%$ of the samples observed.

Just like:

- A fair die has known probabilities
- A single roll has no probability attached after it occurs
- Yet the fairness claim is still meaningful

→ Single sample has no probability after observed

→ Yet statements about long run probability is still true

7. Why this feels unsatisfying (and rightly so)

What you are feeling is exactly why:

- Fisher emphasized **evidence** (p-values)
- Neyman–Pearson emphasized **decision rules**
- Bayesians rejected this framing entirely

Bayesian credible intervals *do* allow:

$$P(\theta \in \text{interval} \mid \text{data})$$

Frequentist confidence intervals **deliberately refuse** to do this.

Your discomfort is not confusion — it is philosophical consistency.

8. One-sentence resolution

A confidence interval is not used to describe what this one sample “means”; it is used to implement a decision rule whose error rate is controlled across hypothetical repetitions.

That is why a single observed CI can legitimately be used to reject a hypothesis.

If you want, next we can:

- Contrast this with **Bayesian credible intervals** precisely
- Show a **formal proof** of CI $\leftrightarrow$ hypothesis test equivalence
- Discuss why **scientists still interpret CIs incorrectly**
- Explain how **likelihood-based inference** fits between CI and Bayes

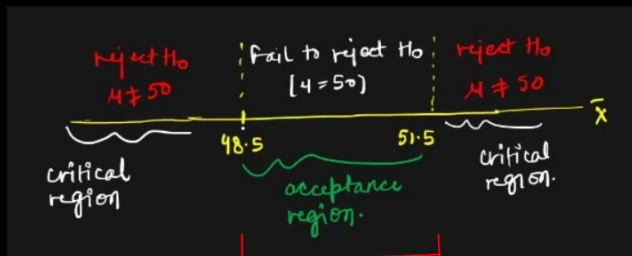
You're asking exactly the right questions — just tell me where you want to go next.



To illustrate the general concepts, consider the propellant burning rate problem introduced earlier. The null hypothesis is that the mean burning rate is 50 centimeters per second, and the alternate is that it is not equal to 50 centimeters per second. That is, we wish to test

$$H_0: \mu = 50 \text{ centimeters per second}$$

$$H_1: \mu \neq 50 \text{ centimeters per second}$$



How do we come up with this region?

→ Assume, we have not started sampling
→ But we already know the sample size & the sampling distribution.

$$E(x_i) = \mu \quad \text{Var}(x_i) = \sigma^2 \quad \left| \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad \right| \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\textcircled{1} \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n) \rightarrow \text{if } n > 30 \text{ using CLT} \\ \rightarrow \text{if } n < 30 \text{ \& } x_i \sim \text{Normal}(\mu, \sigma^2)$$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim N(0,1) \rightarrow \text{if } n > 30$$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} \rightarrow \text{if } n < 30$$

② Using ① We can compute the unobserved confidence interval for $\bar{X}$ as $P(\bar{X} - z^* \cdot SE \leq \mu \leq \bar{X} + z^* \cdot SE) = 1 - \alpha = \text{Confidence}$
which is equal to $P(\mu - z^* \cdot SE \leq \bar{X} \leq \mu + z^* \cdot SE) =$

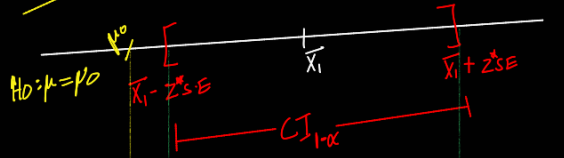
→ The probability of the population parameter falling within the unobserved calculated sample interval is same as the probability of the unobserved sample statistic falling within the same interval about true population parameter which is $= 1 - \alpha = \text{confidence}$

→ When α increases, there is reduced confidence requirement & the CI tends to be smaller
→ shrinking $CI_{1-\alpha}$ increases α & expanding $CI_{1-\alpha}$ decrease α

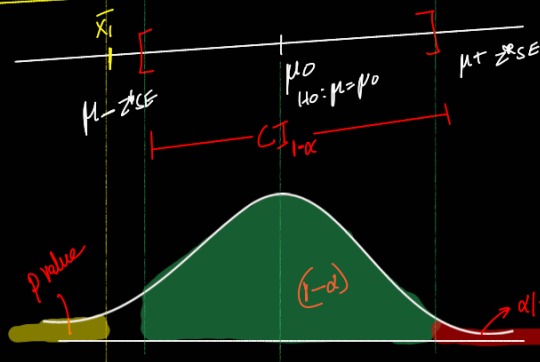
→ if a significance level of α rejects $H_0 \Rightarrow$ The confidence level of $1 - \alpha$ does not contain the null μ_0

$P(\bar{x}_1 | H_0) = \text{p-value} \rightarrow \text{fixed}$
→ of a yet to be realised hypothetical $\bar{x}$

Intuition 1



Intuition 2



Intuition 1

- ① There is $(1-\alpha)\%$ chance that population parameter falls in $CI_{1-\alpha}$
- ② There is $\alpha\%$ chance that population parameter falls outside $CI_{1-\alpha}$
- ③ if $p\text{-value} < \alpha \Rightarrow \text{Reject } H_0$
- ④ Expand $CI_{1-\alpha}$ (ie decrease α) such that H_0 is @ edge of $CI_{1-\alpha}$
 $\Rightarrow p\text{-value} = \alpha^*$
- ⑤ Definition of p-value = ① Smallest value of α such that H_0 is rejected
② Largest value of α such that H_0 is accepted

Intuition 2

- ① Assuming that H_0 is true, the confidence interval $CI_{1-\alpha}$ about the μ_0 is $(1-\alpha)\%$ confident that the observed $\bar{x}_1$ lies inside $CI_{1-\alpha}$
- ② This is equivalent to saying there is $(1-\alpha)\%$ confidence that a observation like x_1 does not occur outside the $CI_{1-\alpha}$
- ③ Still if it is observed, there is only $\alpha\%$ chance that this extreme observation gets repeated in the long run.
 $(1-\alpha)\%$ of time this observation doesn't repeat give H_0 is true.
- ④ if the $p\text{-value} > \alpha\%$, then we don't have enough confidence to reject H_0
- ⑤ if $p < \alpha$ then we have $(1-\alpha)\%$ confidence in rejecting H_0
- ⑥ Maximum confidence occurs when $p\text{-value} \text{ just } < \alpha \Rightarrow$
 $p\text{-value} = \text{smallest } \alpha \text{ such that we can reject } H_0 \text{ with maximum confidence.}$

Caution

→ if through we reject H_0 or fail to reject H_0 based on each sample individually.
hypothesis testing doesn't guarantee anything about the confidence of each individual sample decision.
→ Hypothesis testing only guarantees that long running rejection procedure will be accurate $(1-\alpha)\%$ of the time
→ Hypothesis testing $\neq$ probability $\rightarrow$ Optimum decision rule that guarantees long running probability

3) Suppose the courage of psychologists has a population that is normally distributed with a standard deviation of 10. You decide to sample 57 psychologists from this population and obtain a mean courage of 34.81 and a standard deviation of 9.0579. Using an alpha value of $\alpha = 0.05$, is this observed mean significantly greater than an expected courage of 34?

one tailed

$Z_{critical} = 1.64$ for $\alpha = 0.05$

a) z-value observed is 0.61, we need to perform two tailed test and the observed mean and population mean are same.

✓ b) z-value observed is 0.61, we need to perform right tailed test and the observed mean and population mean are same.

c) z-value observed is 0.61, we need to perform left tailed test and the observed mean and population mean are same.

d) None of the above

population SD = 10
 $\sigma^2 = 100$

$n = 57 \Rightarrow CLT$ applicable

$\bar{X}_1 = 34.81$

$S_1 = 9.0579$

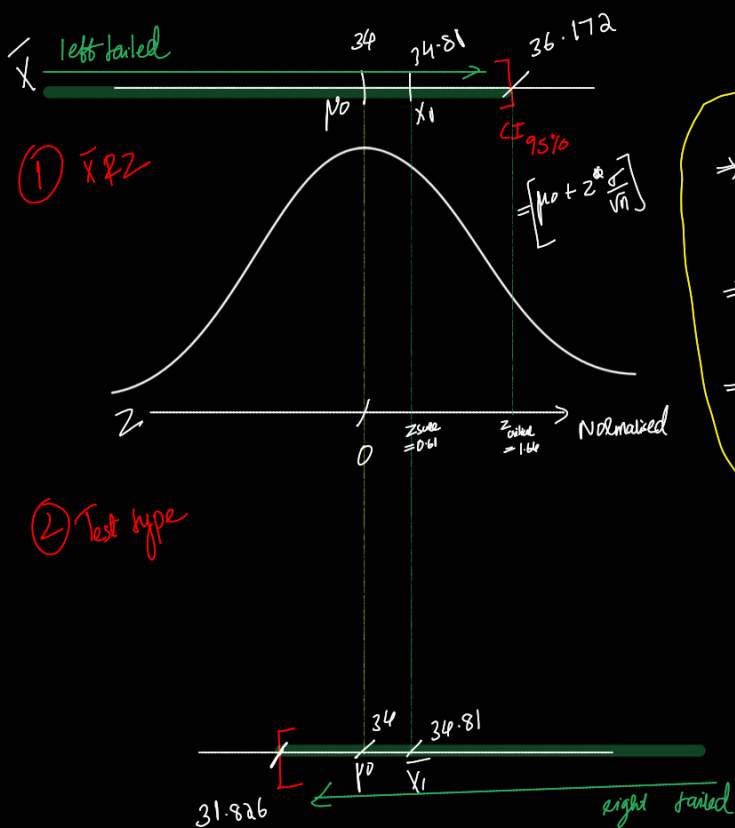
$H_0: \mu = 34$

$H_1: \mu > 34 \Rightarrow$ one tailed — right tailed

$\alpha = 0.05 \Rightarrow 1 - \alpha = 95\%$

$CI_{95\%} \Rightarrow 1.64$

one tailed



① $\bar{x} = 34.81$

② Test type

$$P\left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq 1.64\right] = 0.95 \quad \text{Z-test}$$

$$\Rightarrow P\left[X \leq \mu + 1.64 \times \left(\frac{\sigma}{\sqrt{n}}\right)\right] = 0.95$$

$$\Rightarrow P\left[X \leq 34 + \left(1.64 \times \frac{10}{\sqrt{57}}\right)\right] = 0.95$$

$$\Rightarrow P[X \leq 36.172] = 0.95 \Rightarrow \text{fails to reject } H_0$$

$\Rightarrow X_1$ within $CI_{1-\alpha}$ about μ_0 left tailed

$$P\left[\mu \geq \bar{X} - z^* \frac{\sigma}{\sqrt{n}}\right] = 0.95$$

$$P\left[\mu \geq 34 - 1.64 \times \frac{10}{\sqrt{57}}\right] = 0.95$$

$$P\left[\mu \geq 31.826\right] = 0.95$$

μ_0 within $CI_{1-\alpha}$ about X_1 right tailed

Thus is the statement about μ

$\mu \in CI(X)$ where $CI(X)$ is a random interval

$\Rightarrow$ The right tailness or left tailness only applies when making statement about population parameters $\Rightarrow CI_{1-\alpha}$ about X_1

① Interpretation valid but not a statement about population parameters

② Correct interpretation about population parameters used to signify tailness

Trick $\rightarrow$ tailness of $H_a \Rightarrow$ tailness of $CI_{1-\alpha}$ about the observed value X_1

Alternative hypothesis	Test type	Confidence interval
$\mu > \mu_0$	Right-tailed	Lower bound CI
$\mu < \mu_0$	Left-tailed	Upper bound CI
$\mu \neq \mu_0$	Two-tailed	Two-sided CI

Normalised graph (Z_{score})
 Right rejecting
 left rejecting
 Two sided rejecting

Z_{score} $\left\{ \begin{array}{l} \text{Falling in tail} \Rightarrow \text{Reject } H_0 \\ \text{not in tail} \Rightarrow \text{Fail to reject} \end{array} \right.$ ☆☆☆

8) One day you sample the knowledge of 31 utopian dinosaurs from a population and obtain a mean knowledge of 45.64 and a standard deviation of 7.407. Using an alpha value of $\alpha = 0.01$, is this observed mean significantly less than an expected knowledge of 50?

$$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim N(0,1)$$

$$P\left(\frac{\bar{X} - \mu}{s/\sqrt{n}} \geq -Z\right) = 0.99$$

$$P\left(\mu \leq \bar{X} + \frac{c \cdot s}{\sqrt{n}}\right) = 0.99$$

$$Z_{\text{score}} = \frac{45.64 - 50}{7.407/\sqrt{31}} = -3.277$$

$$Z_{\text{test}} \frac{\alpha}{2} = 0.01 \text{ left tailed} = 2.325$$

$$\Rightarrow Z_{\text{score}} < -2.325 \rightarrow \text{reject } H_0 \rightarrow \text{reject } \mu = 50$$

$$T\text{-test } T^* \alpha = 0.01 \text{ left tailed} = 2.45$$

$$T_{\text{score}} < -2.45 \rightarrow \text{reject } H_0 \rightarrow \text{reject } \mu = 50$$

$$\bar{X}_1 = 45.64$$

$$S_1 = 7.407$$

$$n = 31 \Rightarrow \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim T(30) \sim N(0,1) \text{ as } n \text{ large}$$

$$\alpha = 0.01$$

$$H_0 = 50$$

$$H_a = \mu < 50 \text{ --- left tailed}$$

6) Suppose the courage of hateful republicans has a population that is normally distributed with a standard deviation of 6. Let's pretend that you sample 58 hateful republicans from this population and obtain a mean courage of 23.59 and a standard deviation of 6.61.

Using an alpha value of $\alpha = 0.05$, is this observed mean significantly less than an expected courage of 26?

Z-value for alpha = 0.05 is 1.64

a) The courage of hateful republicans is significantly less than 26.

b) The courage of hateful republicans is significantly more than 26.

c) we perform two tailed test, z-value is +3.06

d) we perform left tailed test, z-value is -3.06

population SD = 6
Var = 36

$$n = 58$$

$$\bar{X}_1 = 23.59$$

$$S_1 = 6.61$$

$$\alpha = 0.05$$

$$H_0: \mu = 26$$

$$H_a: \mu < 26 \text{ (left tailed)}$$

$\Rightarrow$ reject left

Sampling from Normal

$$\Rightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$Z_{\text{score}} = \frac{23.59 - 26}{6/\sqrt{58}} = -3.06$$

$$Z_{\text{critical}} = -1.64$$

$$Z_{\text{score}} < Z_{\text{critical}} \Rightarrow \text{reject } H_0 \Rightarrow \mu < 26$$

Suppose you want to determine if the height of women with fathers 6 feet or taller in our class is significantly different from the average height of women in the US population. According to Wikipedia, the average height of women in the US is 64 inches, but the website does not provide standard deviations. The heights of the 34 women with fathers 6 feet or taller in our class has a mean of 66.62 and a standard deviation of 2.663 inches. Is this mean significantly different from 64 inches? Use an α value of 0.05.

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{33}, P(-T^* \leq T \leq T^*) = 0.95$$

$$T_{\text{two-sided}} \alpha = 0.05 = 2.042$$

$$T_{\text{score}} = \frac{66.62 - 64}{2.663/\sqrt{34}} = 5.137$$

$$\mu = 64$$

$$n = 34$$

$$\bar{X}_1 = 66.62$$

$$S_1 = 2.663$$

$$H_0: \mu = 64$$

$$H_a: \mu \neq 64 \text{ --- two tailed}$$

$$\alpha = 0.05$$

$$T_{\text{score}} \text{ outside } CI_{1-\alpha} \Rightarrow \text{reject } H_0$$

$\Rightarrow$ Sample mean significantly different from 64 inches.

9-23. A melting point test of $n = 10$ samples of a binder used in manufacturing a rocket propellant resulted in $\bar{x} = 154.2^\circ\text{F}$. Assume that melting point is normally distributed with $\sigma = 1.5^\circ\text{F}$.

(a) Test $H_0: \mu = 155$ versus $H_a: \mu \neq 155$ using $\alpha = 0.01$.

(b) What is the P-value for this test?

$$n = 10$$

$$\bar{X}_1 = 154.2^\circ\text{F}$$

$$\sigma = 1.5^\circ\text{F}$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$x_i \sim N(\mu, \sigma^2)$$

$$Z_{\text{score}} = \frac{154.2 - 155}{1.5/\sqrt{10}}$$

$$Z_s = -1.686$$

$$H_0: \mu = 155$$

$$H_a: \mu \neq 155$$

$$Z_{\text{critical}} = 2.57$$

$$\Rightarrow -Z_c \leq Z_s \leq Z_c \Rightarrow \text{can't reject } H_0$$

$$b) \text{ p value} = \text{value of } \alpha \text{ at boundary}$$

$$\Rightarrow Z_{\text{critical}} = 1.686 \Rightarrow 1 - \alpha/2 = 0.9545 \text{ --- from Z table}$$

$$1 - \alpha/2 = 0.9545$$

$$\alpha = 2[1 - 0.9545] = 2 \times 4.55\% = 9.10\%$$

$$\text{p value} = 0.091$$

9-20. The mean water temperature downstream from a power plant cooling tower discharge pipe should be no more than 100°F. Past experience has indicated that the standard deviation of temperature is 2°F. The water temperature is measured on nine randomly chosen days, and the average temperature is found to be 98°F.

Should the water temperature be judged acceptable with $\alpha = 0.05$?

(b) What is the P-value for this test?

$$H_0: \mu \geq 100^\circ\text{F}$$

$$H_a: \mu < 100^\circ\text{F} \text{ left tail}$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\sigma = 2^\circ\text{F}$$

$$\bar{X} = 98^\circ\text{F}$$

$$P(Z > -Z) = 1 - \alpha$$

$$@ \alpha = p\text{-value}$$

$$Z_5 = Z^* = 3$$

$$1 - p = P(Z > 3)$$

$$1 - p = P(Z < 3)$$

$$p = P(Z > 3)$$

interchanging 170 & 110
interchanges p.

Q.20	A random sample $X_1, X_2, \dots, X_6$ of size 6 is taken from a Bernoulli distribution with the parameter θ . The null hypothesis $H_0: \theta = \frac{1}{2}$ is to be tested against the alternative hypothesis $H_1: \theta > \frac{1}{2}$, based on the statistic $Y = \sum_{i=1}^6 X_i$. If the value of Y corresponding to the observed sample values is 4, then the p-value of the test statistic is
(A)	$\frac{21}{32}$
(B)	$\frac{9}{64}$
(C)	$\frac{11}{32}$
(D)	$\frac{7}{64}$

$X \sim \text{Bern}(\theta)$

$H_0: \theta = \frac{1}{2}$

$H_1: \theta > \frac{1}{2}$

$Y = \sum_{i=1}^6 X_i = X_1 + \dots + X_6$

$Y = 4$

$$X_i \sim \text{Bern}(\theta)$$

$$H_0: \theta = 1/2$$

$$H_a: \theta > 1/2 \text{ right tailed}$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n (x_i) = \text{Sample Sum}$$

$$\bar{Y}_1 = 4$$

P-value of test = ?

$$p\text{-value} = P(X \text{ as extreme as observed} / n = 4)$$

$$= P(Y = 4 \text{ or } Y = 5 \mid \theta = 1/2)$$

$$= 6C_4 (1/2)^6 + 6C_5 (1/2)^6$$

$$= \frac{1}{2^6} [15 + 6] = \frac{21}{64}$$

$$P = \frac{11}{32}$$

Q.4	Let $\{x_1, x_2, \dots, x_n\}$ be a realization of a random sample of size $n (\geq 2)$ from a $N(\mu, \sigma^2)$ distribution, where $-\infty < \mu < \infty$ and $\sigma > 0$. Which of the following statements is/are true? P: 95% confidence interval of μ based on $\{x_1, x_2, \dots, x_n\}$ is unique when σ is known. Q: 95% confidence interval of μ based on $\{x_1, x_2, \dots, x_n\}$ is NOT unique when σ is unknown.
(A)	P only
(B)	Q only
(C)	Both P and Q
(D)	Neither P nor Q

A simple random sample of size 10 from $N(\mu, \sigma^2)$ gives 98% confidence interval (20.49, 23.51). Then the null hypothesis $H_0: \mu = 20.5$ against $H_a: \mu \neq 20.5$

- (A) can be rejected at 2% level of significance
- (B) cannot be rejected at 5% level of significance
- (C) can be rejected at 10% level of significance
- (D) cannot be rejected at any level of significance

EXAMPLE 9-2 Propellant Burning Rate

Air crew escape systems are powered by a solid propellant. The burning rate of this propellant is an important product characteristic. Specifications require that the mean burning rate must be 50 centimeters per second. We know that the standard deviation of burning rate is $\sigma = 2$ centimeters per second. The experimenter decides to specify a type I error probability or significance level of $\alpha = 0.05$ and selects a random sample of $n = 25$ and obtains a sample average burning rate of $\bar{x} = 51.3$ centimeters per second. What conclusions should be drawn?

We may solve this problem by following the seven-step procedure outlined in Section 9-1.6. This results in

1. **Parameter of interest:** The parameter of interest is μ , the mean burning rate.
2. **Null hypothesis, H_0 :** $H_0: \mu = 50$ centimeters per second
3. **Alternative hypothesis, H_1 :** $H_1: \mu \neq 50$ centimeters per second

4. **Test statistic:** The test statistic is

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

5. **Reject H_0 if:** Reject H_0 if the P-value is less than 0.05. To use a fixed significance level test, the boundaries of the critical region would be $z_{0.025} = 1.96$ and $-z_{0.025} = -1.96$.

6. **Computations:** Since $\bar{x} = 51.3$ and $\sigma = 2$,

$$z_0 = \frac{51.3 - 50}{2/\sqrt{25}} = 3.25$$

7. **Conclusion:** Since $z_0 = 3.25$ the p-value is $P\text{-value} = 2[1 - \Phi(3.25)] = 0.0012$

we reject $H_0: \mu = 50$ at the 0.05 level of significance.

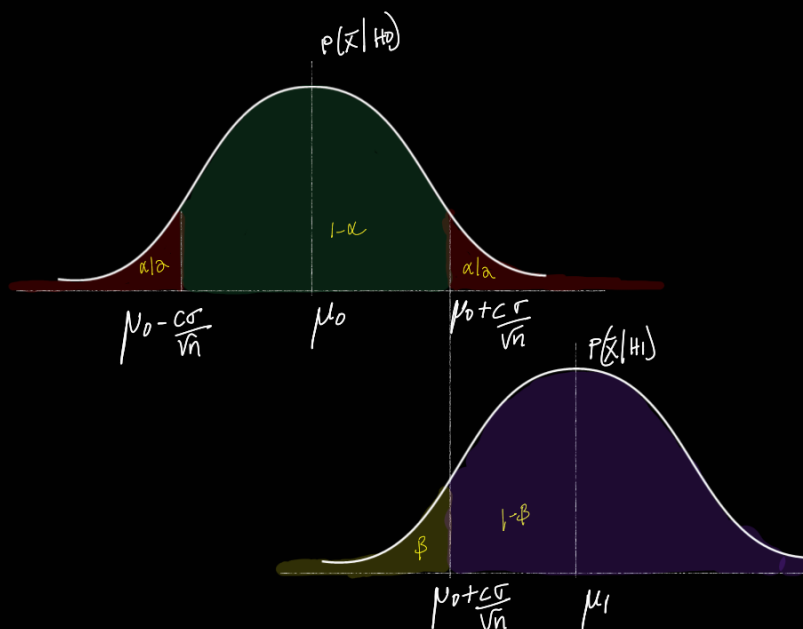
Practical Interpretation: We conclude that the mean burning rate differs from 50 centimeters per second, based on a sample of 25 measurements. In fact, there is strong evidence that the mean burning rate exceeds 50 centimeters per second.

Type 1 error, Type 2 error & power

	H_0 actually true	H_0 actually false
Reject H_0	Type 1 error = α $P(\text{Reject } H_0 H_0)$	$1 - \beta$ Power
Fail to Reject H_0	$1 - \alpha$ $P(\text{Fail to Reject } H_0 H_0)$	Type 2 error = β $P(\text{Fail to Reject } H_0 H_1)$

$$H_0: \mu = \mu_0$$

$$H_1: \mu = \mu_1$$



graph if $\mu_1 > \mu_0$

9-2. A textile fiber manufacturer is investigating a new drapery yarn, which the company claims has a mean thread elongation of 12 kilograms with a standard deviation of 0.5 kilograms. The company wishes to test the hypothesis $H_0: \mu = 12$ against $H_1: \mu < 12$, using a random sample of four specimens.

- (a) What is the type I error probability if the critical region is defined as $\bar{x} < 11.5$ kilograms?
- (b) Find β for the case where the true mean elongation is 11.25 kilograms.

b) $H_0: \mu = 12$
 $H_1: \mu = 11.25$

$$\beta = \text{Type II} = P(\text{Fail to Reject } H_0 | H_1 \text{ true})$$

$$= P(\bar{X} > 11.5 | \mu = 11.25)$$

$$= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{11.5 - 11.25}{0.25} \right) \Rightarrow P(Z > 1) = \beta$$

$$1 - P(Z < 1) = \beta$$

$$0.8413 = 1 - \alpha \Rightarrow \alpha = 0.1797 = 17.97\%$$

a) $H_0: \mu = 12, \sigma = 0.5$

$\bar{x} < 11.5$ critical

$$\Rightarrow \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \frac{11.5 - 12}{0.25}$$

$Z < -2$ — critical — $P(Z < -2) = \alpha$

$Z > -2$ — acceptance — $P(Z > -2) = 1 - \alpha$

$$\Rightarrow P(Z < -2) = 1 - \alpha$$

$$1 - 0.9772 = \alpha$$

$$\alpha = 0.0228 = 2.28\%$$

Let p be the probability that a coin will fall heads in a single toss in order to test

$H_0: p = 0.5$

$H_1: p = 0.75$

The coin is tossed 5 times and H_0 is rejected if more than 3 heads are obtained. Find the probability of Type I error and the power of the test.

$$\text{Power} = P(\bar{X} > 3 | p = 0.75) = 5C_4 \cdot \left(\frac{3}{4}\right)^4 \cdot \frac{1}{4} + 5C_5 \cdot \left(\frac{2}{4}\right)^5$$

$$= \frac{5 \times 81}{4^5} + \frac{3 \times 81}{4^5} = \frac{8 \times 81}{4 \times 4 \times 64} = \frac{81}{128}$$

$\bar{X} > 3$ — critical region

$\bar{X} \leq 3$ — acceptance region

$n = 5$ times

$\bar{X} = \sum X_i$

$X_i = \begin{cases} 1 & \text{if toss is head} \\ 0 & \text{if toss is tail} \end{cases}$

$$\text{Type I} = P(\bar{X} > 3 | p = 0.5)$$

$$= \frac{5C_4 + 5C_5}{32} = \frac{5 + 1}{32} = \frac{3}{16}$$

Let X have a binomial distribution with parameters n and p , $n = 3$. For testing the hypothesis $H_0: p = 2/3$ against $H_1: p = 1/3$, let a test be Reject H_0 if $X \geq 2$ and accept H_0 if $X \leq 1$. Then the probabilities of Type I and Type II errors respectively are

- a) $\frac{20}{27}$ and $\frac{20}{27}$
 b) $\frac{7}{27}$ and $\frac{20}{27}$
 c) $\frac{20}{27}$ and $\frac{7}{27}$
 d) $\frac{7}{27}$ and $\frac{7}{27}$

The probability density function of the random variable X is

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Where $\lambda > 0$. For testing the hypothesis $H_0: \lambda = 3$ against $H_A: \lambda = 5$, a test is given as
"Reject H_0 if $X \geq 4.5$ ".

The probability of type I error and power of test are respectively

- a) 0.1353 and 0.4966 b) 0.1827 and 0.379
 c) 0.2021 and 0.4493 d) 0.2231 and 0.4066

exponential PDF, $f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}$ | CDF = $1 - e^{-x/\lambda}$

$\lambda = 3$
 $\lambda = 5$ Critical region (R) = $X \geq 4.5$

power = $P(X \geq 4.5 | \lambda = 3) = e^{-\frac{4.5}{3}} = 0.2231$
 $\alpha = P(X \geq 4.5 | \lambda = 5) = e^{-\frac{4.5}{5}} = 0.4066$

Suppose that X is a population random variable with probability density function

$$f(x; \theta) = \begin{cases} \theta x^{\theta-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where θ is a parameter. In order to test the null hypothesis

$H_0: \theta = 2$, against the alternative hypothesis $H_1: \theta = 3$,

the following test is used: Reject the null hypothesis

if $X_1 \geq \frac{1}{2}$ and accept otherwise,

where X_1 is a random sample of size 1

drawn from the above population. Then the

power of the test is 7/8

$$F_X(x) = \theta x^{\theta-1} \quad 0 < x < 1$$

$$H_0: \theta = 2$$

$$H_1: \theta = 3$$

$X_1 \geq 1/2 \rightarrow$ critical region $\Rightarrow$ reject H_0

$X_1 < 1/2 \rightarrow$ acceptance region $\Rightarrow$ fail to reject H_0

power = $1 - \beta = P(X_1 \geq 1/2 | \theta = 3) = \int_{1/2}^1 3x^2 = [x^3]_{1/2}^1$
 $= 1 - \frac{1}{8} = \frac{7}{8}$

Let X_1 be a random sample of size 1 from uniform distribution over (θ, θ^2) where $\theta > 1$. To test $H_0: \theta = 2$

against $H_1: \theta = 3$, reject H_0 if and only if $X_1 > 3.5$. Let α

and β be the size and power, respectively, of this test.

Then $\alpha + \beta$ (rounded off to two decimal places) is equal to

$$X \sim \text{Uniform}(\theta, \theta^2)$$

$$\bar{X}_1 \rightarrow \text{Sample } n=1$$

$$H_0: \theta = 2$$

$$H_A: \theta = 3$$

$X_1 > 3.5$ critical region

$$\alpha = P(X > 3.5 | \theta = 2) = \frac{4 - 3.5}{4 - 2} = 1/4$$

$$\beta = P(X > 3.5 | \theta = 3) = \frac{9 - 3.5}{9 - 3} = 1/2$$

$$\alpha + \beta = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

Let X_1, X_2, X_3, X_4 be a random sample of size four from a Bernoulli distribution with the parameter θ , $0 < \theta < 1$.

Consider the null hypothesis $H_0: \theta = 1/4$ against the alternative hypothesis $H_1: \theta > 1/4$. Suppose H_0 is rejected if and only if $X_1 + X_2 + X_3 + X_4 > 2$. If α is the probability of Type I error for the test and $\gamma(\theta)$ is the power function of the test, then the value of

$$16\alpha + 7\gamma\left(\frac{1}{2}\right) \text{ (in integer) is equal to } \underline{20}$$

$$\alpha = P(X_1 + X_2 + X_3 + X_4 < 2 | \theta = 1/4) = \binom{4}{0} \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^4 + \binom{4}{1} \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^3 + \binom{4}{2} \left(\frac{1}{4}\right)^2 \left(\frac{3}{4}\right)^2$$

$$16\alpha = \frac{81}{16} + \frac{27 \times 4}{16} + \frac{54}{16} \Rightarrow 16\alpha + 7\gamma = \frac{11 \times 7 + 54 + 108 + 81}{16}$$

$$\frac{277}{16} = \frac{308}{16} - \frac{20}{16}$$

Chi Squared distribution χ_n^2

Let $S_n = X_1^2 + X_2^2 + X_3^2 + \dots + X_n^2$ | $n = \text{degree of freedom}$
 $X_1, X_2, X_3, \dots, X_n$ are iid Normal(0,1)

$$\Rightarrow S_n \sim \chi^2(n) \quad S_n \sim \text{gamma}(n/2, 1/2)$$

$$E(S_n) = n$$

$$\text{Var}(S_n) = 2n$$

$$E(S_n^k) = n(n+2)(n+4) \dots [n+2(k-1)]$$

$$E(S_n^k) = \prod_{i=0}^{k-1} (n+2i)$$

if $X_1, X_2, \dots$ are Normal(μ, σ^2) instead of Normal(0,1)

$$\Rightarrow \text{Let } Z_1 = \frac{X_1 - \mu}{\sigma} \sim \text{Normal}(0, 1)$$

$$X_1 = \sigma Z_1 + \mu$$

$$X_1^2 = \sigma^2 Z_1^2 + 2\sigma\mu Z_1 + \mu^2$$

$$E(X_1^2) = \sigma^2 + \mu^2 \quad \left| \begin{array}{l} \text{since } E(Z_1) = 0 \\ E(Z_1^2) = \text{Var}(Z_1) = 1 \end{array} \right.$$

$$\text{Var}(X_1^2) = \text{Var}(\sigma^2 Z_1^2) + \text{Var}(2\sigma\mu Z_1) \quad \left| \begin{array}{l} \text{since } Z_1 \neq Z_1^2 \\ \text{unrelated} \end{array} \right.$$

$$\text{Var}(X_1^2) = \sigma^4 [E(Z_1^4) - \frac{E(Z_1^2)^2}{1}] + 4\sigma^2 \mu^2$$

$$\text{Var}(X_1^2) = 2\sigma^4 + 4\sigma^2 \mu^2$$

Let $X_i, i=1, 2, \dots, n$, be i.i.d. random variables from a normal distribution with mean 1 and variance 4. Define

$$S_n = X_1^2 + X_2^2 + \dots + X_n^2$$

If $\text{Var}(S_n)$ denotes the variance of S_n , then find the value of

$$\lim_{n \rightarrow \infty} \left(\frac{\text{Var}(S_n)}{n} - \left(\frac{E[S_n]}{n} \right)^2 \right)$$

$$X_1^2 \sim \text{Normal}(1, 4)$$

$$E(X_1^2) = 4 + 1 = 5$$

$$\text{Var}(X_1^2) = 2 \times 16 + 4 \times 4 \times 1 = 32 + 16 = 48$$

$$E[S_n] = n E(X_1^2) = 5n$$

$$\text{Var}[S_n] = n \text{Var}(X_1^2) = 48n$$

$$\lim_{n \rightarrow \infty} \left(\frac{\text{Var}(S_n)}{n} - \left(\frac{E(S_n)}{n} \right)^2 \right) = 48 - 25 = 23$$

Chi-Square Goodness-of-Fit: Intuition & Justification (Complete Summary)

1. What are we testing?

- We compare **observed counts** O_i in buckets to **expected counts** $E_i = np_i$ under a hypothesized distribution.
- The chi-square statistic is

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

2. What is the random variable?

- Each bucket count is a **count random variable**, not raw data:

$$O_i \sim \text{Binomial}(n, p_i)$$

- Mean:

$$\mathbb{E}[O_i] = E_i$$

- Variance:

$$\text{Var}(O_i) = np_i(1 - p_i) \approx np_i = E_i$$

(The approximation holds when p_i is small, which is typical.)

3. Why does normality apply?

- When expected counts are large ($E_i \gtrsim 5$):

$$\frac{O_i - E_i}{\sqrt{E_i}} \xrightarrow{d} \mathcal{N}(0, 1)$$

(by CLT / normal approximation to binomial).

4. Why divide by E_i ?

- Since $\text{Var}(O_i) \approx E_i$, we are effectively computing:

$$\frac{(O_i - E_i)^2}{\text{Var}(O_i)}$$

- This is the **square of a standardized normal variable**:

$$\left(\frac{O_i - \mu_i}{\sigma_i} \right)^2$$

5. Interpretation of the denominator

The denominator E_i acts as:

- a **variance stabilizer**
- a **scaling / weighting factor**
- a way to ensure **uniform treatment of buckets**
- prevents buckets with large expected counts (and naturally larger variance) from dominating the statistic

So each bucket contributes on the same **standardized scale**.

6. Why does the sum follow a chi-square distribution?

- Each term is approximately Z_i^2 where $Z_i \sim \mathcal{N}(0, 1)$
- The sum of squared independent standard normals follows:

$$\chi^2$$

- Degrees of freedom:

$$\text{df} = k - 1$$

(because counts must sum to n , imposing one constraint).

7. Key conditions for validity

- Expected counts should be sufficiently large (rule of thumb: $E_i \geq 5$)
- This ensures the normal and chi-square approximations are accurate

One-line intuition

The chi-square statistic is the sum of **variance-normalized squared deviations** of observed bin counts from their expected values.

Chi-Square Test of Independence: Intuition & Justification (Complete Summary)

1. What are we testing?

- We test whether **two categorical variables are independent**.
- Data are arranged in a **contingency table** with observed counts O_{ij} .
- The chi-square statistic is:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

2. What are the expected counts?

- Under the null hypothesis (independence):

$$E_{ij} = \frac{(\text{row total}_i)(\text{column total}_j)}{\text{grand total}}$$

- These represent the **mean counts** we expect if rows and columns are unrelated.

3. What is the random variable?

- Each cell count is a **count random variable**:

$$O_{ij} \sim \text{Multinomial cell count}$$

- Mean:

$$E[O_{ij}] = E_{ij}$$

- Variance:

$$\text{Var}(O_{ij}) \approx E_{ij}$$

(Valid when probabilities are small and expected counts are not tiny.)

4. Why does normality apply?

- When expected counts are sufficiently large ($E_{ij} \gtrsim 5$):

$$\frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

(by CLT / normal approximation to multinomial counts).

5. Why divide by E_{ij} ?

- Since $\text{Var}(O_{ij}) \approx E_{ij}$, each term is approximately:

$$\frac{(O_{ij} - E_{ij})^2}{\text{Var}(O_{ij})}$$

- This is the **square of a standardized normal variable**:

$$\left(\frac{O_{ij} - \mu_{ij}}{\sigma_{ij}} \right)^2$$

6. Interpretation of the denominator

The denominator E_{ij} acts as:

- a **variance stabilizer**
- a **scaling / weighting factor**
- ensures **uniform treatment of cells**
- prevents cells with large expected counts (and larger variance) from dominating

Each cell contributes on the same **standardized scale**.

7. Why does the sum follow a chi-square distribution?

- Each standardized cell contributes approximately a Z^2
- The sum of squared standard normals follows:

$$\chi^2$$

- Degrees of freedom:

$$\text{df} = (\text{rows} - 1)(\text{columns} - 1)$$

This comes from the **constraints imposed by fixed row and column totals**.

8. Key conditions for validity

- Expected counts should be sufficiently large (rule of thumb: $E_{ij} \geq 5$)
- Ensures accuracy of normal and chi-square approximations

One-line intuition

The chi-square test of independence measures **variance-normalized squared deviations** between observed and expected cell counts under independence.

Let's see if the birthdays in this class are evenly distributed across months, or if there are some months for which students have significantly more birthdays than others. For simplicity, we'll assume that all months have equal probability, even though they vary in length. We'll run a χ^2 test using an alpha value of 0.05.

Here's a table showing the number of birthdays for each month for all 152 students:

observed frequencies of birthdays											
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
18	8	9	11	14	18	15	14	11	16	7	11

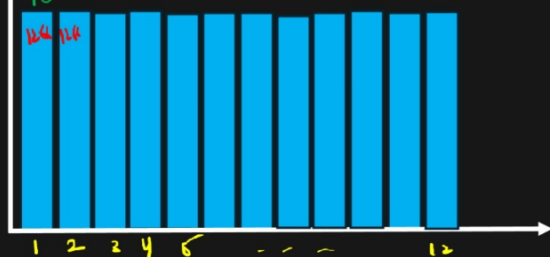
H_0 : birthdays are evenly distrib

Based on 'observed frequ tell whether your hypothesis as to how actual popu look is T or F

H_0 : distribution is uniform
 H_1 : distn " is not uniform

$$\frac{18^2}{12} = 13.66$$

H_0 : H_0



observed frequencies:

1	2	3	4	5	6	7	8	9	10	11	12
18	8	9	11	14	18	16	14	11	6	7	11

$n = 152$

NEED =

find expected distⁿ
find expected frequencies

b_o : observed freq
 b_e : expected

$$\frac{152}{12} = 12.66$$

H_0 :

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
12.66	12.66	12.66	12.66	12.66	12.66	12.66	12.66	12.66	12.66	12.66	12.66

(H_0)

Goal: How far or close observed frequencies are to expected frequency

$$\chi^2 = \sum_{i=1}^n \frac{(b_o - b_e)^2}{b_e} = \sum_{i=1}^n \frac{(O - E)^2}{E} \sim \chi^2_{n-1}$$

statistic

$$\chi^2 = \frac{(18 - 12.66)^2}{12.66} + \frac{(8 - 12.66)^2}{12.66} + \frac{(9 - 12.66)^2}{12.66} + \dots + \frac{(11 - 12.66)^2}{12.66}$$

$$\chi^2 = 12.05$$

$$\chi^2_{obs} = 12.05$$

$$\alpha = 0.05 \quad \chi^2_{critical}$$

Chi-square Distribution Table

Area under curve

d.f.	0.995	0.99	0.975	0.95	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
2	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
3	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
4	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
5	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
6	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
7	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
8	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
9	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
10	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
11	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
12	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
13	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
14	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
15	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
16	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
17	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
18	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
19	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
20	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
22	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
24	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
26	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
28	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
30	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
32	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
34	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
36	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
38	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
40	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
42	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
44	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
46	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
48	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
50	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
55	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
60	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
65	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
70	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
75	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
80	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
85	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
90	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
95	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084
100	0.0001	0.0002	0.0005	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.013	0.016	0.020	0.024	0.030	0.039	0.051	0.068	0.084

	P																	
DF	0.995	0.975	0.95	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	0.0000393	0.0000982	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
2	0.0100	0.0506	0.219	0.460	0.675	0.878	1.064	1.239	1.392	1.528	1.649	1.753	1.849	1.936	2.015	2.079	2.133	2.177
3	0.0717	0.216	0.464	0.625	0.781	0.924	1.077	1.230	1.383	1.524	1.651	1.766	1.862	1.941	2.011	2.069	2.119	2.163
4	0.207	0.484	0.789	1.064	1.286	1.478	1.636	1.773	1.891	1.995	2.088	2.171	2.245	2.309	2.367	2.419	2.466	2.508
5	0.412	0.831	1.279	1.729	2.079	2.365	2.602	2.797	2.957	3.099	3.229	3.347	3.453	3.548	3.633	3.709	3.777	3.838
6	0.676	1.237	1.868	2.445	2.877	3.267	3.619	3.927	4.191	4.419	4.607	4.763	4.897	5.013	5.113	5.200	5.275	5.345
7	0.989	1.690	2.365	3.000	3.518	3.942	4.373	4.754	5.084	5.371	5.617	5.827	6.004	6.151	6.281	6.397	6.502	6.597
8	1.344	2.180	2.998	3.745	4.353	4.879	5.318	5.678	5.969	6.191	6.353	6.496	6.621	6.731	6.828	6.914	7.000	7.076
9	1.735	2.700	3.599	4.426	5.034	5.560	6.000	6.361	6.652	6.884	7.057	7.200	7.325	7.435	7.532	7.618	7.694	7.761
10	2.156	3.247	4.169	4.996	5.604	6.130	6.570	6.931	7.222	7.454	7.627	7.770	7.895	8.005	8.102	8.188	8.264	8.331
11	2.603	3.816	4.761	5.598	6.206	6.732	7.172	7.533	7.824	8.056	8.229	8.372	8.497	8.607	8.704	8.790	8.866	8.933
12	3.074	4.404	5.371	6.218	6.826	7.352	7.792	8.153	8.444	8.676	8.849	8.992	9.117	9.227	9.324	9.410	9.486	9.553
13	3.565	5.009	5.996	6.843	7.451	7.977	8.417	8.778	9.069	9.301	9.474	9.617	9.742	9.852	9.949	10.035	10.111	10.178
14	4.075	5.629	6.636	7.483	8.091	8.617	9.057	9.418	9.709	9.941	10.114	10.257	10.382	10.492	10.589	10.675	10.751	10.818
15	4.601	6.262	7.289	8.136	8.744	9.270	9.710	10.071	10.362	10.594	10.767	10.910	11.035	11.145	11.242	11.328	11.404	11.471
16	5.142	6.908	8.055	8.902	9.510	10.036	10.476	10.837	11.128	11.360	11.533	11.676	11.791	11.891	11.977	12.053	12.120	12.177
17	5.697	7.564	8.731	9.578	10.186	10.712	11.152	11.513	11.804	12.036	12.209	12.352	12.467	12.567	12.653	12.729	12.796	12.853
18	6.265	8.231	9.418	10.265	10.873	11.400	11.840	12.191	12.482	12.714	12.887	13.030	13.145	13.245	13.331	13.407	13.474	13.531
19	6.844	8.907	10.114	10.961	11.569	12.096	12.536	12.887	13.178	13.410	13.583	13.726	13.841	13.941	14.027	14.103	14.170	14.227
20	7.434	9.591	10.818	11.665	12.273	12.800	13.240	13.591	13.882	14.114	14.287	14.430	14.545	14.645	14.731	14.807	14.874	14.931

9) Suppose balloons come in 3 varieties: neat, easy and deep. Tomorrow you find 44 balloons and count how many fall into each variety. This generates the following table:

observed frequencies of balloons		
neat	easy	deep
21	13	10

Make a bar graph showing the frequencies for each variety of balloons

Make a table of the expected frequencies.

Using an alpha value of $\alpha = 0.05$ test the null hypothesis that the 44 balloons are distributed evenly across the 3 varieties of neat, easy and deep.

5) Suppose chickens come in 7 varieties: sedate, careless, rainy, hurried, judicious, outstanding and stormy. I'd like you to find 176 chickens and count how many fall into each variety. This generates the following table:

observed frequencies of chickens						
sedate	careless	rainy	hurried	judicious	outstanding	stormy
22	27	14	26	22	29	36

Make a table of the expected frequencies.

Using an alpha value of $\alpha = 0.05$ test the null hypothesis that the 176 chickens are distributed evenly across the 7 varieties of sedate, careless, rainy, hurried, judicious, outstanding and stormy.

Q) Does the type of computer that students use in our class depend on gender? To find out we'll run a χ^2 test for independence using alpha = 0.05.

We'll use our survey data and count the number of students who use each kind of computer, depending upon their gender. This generates the following 2 X 2 table:

observed frequencies		
	Apple	PC
Female	82	32
Male	13	13

X: gender of person
Y: laptop/gound

Female - Apple
Male - Windows

H_0 : comp used does not depend on gender
 H_1 : depends on gender

χ^2 test

How to find expected frequencies?

Observation: A PC

Total # students = 114 + 26 = 140 = N

$$\% F = \frac{114}{140} = 81\%$$

$$\% M = \frac{26}{140} = 19\%$$

The ratio should be considered.
81% of people using A should be F,
19% should be M

expected frequency	
$\frac{114}{140} \times 95$	$\frac{114}{140} \times 45$
$\frac{26}{140} \times 95$	$\frac{26}{140} \times 45$

$$\text{Expected} = \frac{\text{row sum} \times \text{column sum}}{\text{grand total}}$$

$$\text{dof} = (8-1)(2-1)$$

observed

	A	PC	
F	82	32	114
M	13	13	26
	95	45	140

$$\% \text{ people using A} = \frac{95}{140} = 67\%$$

$$\% \text{ using PC} = \frac{45}{140} = 33\%$$

expected distribution.

$\left(\frac{95}{140}\right) \times 114$	$\left(\frac{45}{140}\right) \times 114$
$\left(\frac{95}{140}\right) \times 26$	$\left(\frac{45}{140}\right) \times 26$

expected

77.75	26.54
17.64	8.85

$$\chi^2 \text{ statistic} = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 =$$

$$\chi^2 = \frac{(82-77.3571)^2}{77.3571} + \frac{(13-17.6429)^2}{17.6429} + \frac{(32-36.6429)^2}{36.6429} + \frac{(13-8.3571)^2}{8.3571} = 0.2787 + 1.2218 + 0.5883 + 2.5794 = 4.6682$$

$$\chi^2_{obs} = 4.66$$

$$\chi^2_{critical}$$

$$\chi^2_{\alpha, df} = \chi^2 = ?$$

$$\alpha = 0.05$$

$$\chi^2_{crit} =$$

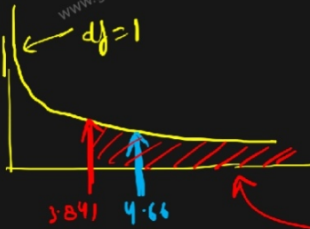
$$df = ?$$

$$df = (r-1)(c-1) = (2-1)(2-1)$$

$$df = 1$$

$$\chi^2_{obs} > \chi^2_{crit}$$

reject H_0



Assumption was wrong
comp used depends on
gender

An observation that is useful: (2×2)

contingency

	Y=0	Y=1	
X=0	a	b	a+b
X=1	c	d	c+d
	a+c	b+d	N

$$\chi^2 = N \times (\text{det of contingency matrix})^2$$

$$\chi^2 = \frac{N \times (ad - bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

$$N = (a+b) + (c+d) = (a+c) + (b+d)$$

2x2 shortcut (only 2x2)

$$\chi^2 = \frac{N \times |\text{Contingency Matrix}|^2}{\text{product of rows sum} \times \text{product of column sum}}$$

N = grand total

eg X=0	a	b	a+b
X=1	c	d	c+d
Y=0	a	c	a+c
Y=1	b	d	b+d

$$a+b+c+d = N$$

$$\chi^2 = \frac{(a+b+c+d)(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

A χ^2 test for independence between two attributes X and Y is carried out at 2.5% level of significance on the following 2×2 contingency table showing frequencies

	X ₁	X ₂	
Y ₁	1	0	1
Y ₂	1	d	1+d

If the upper 2.5% point of the χ^2_1 distribution gives as 5.0, then the hypothesis of independent is to be rejected if and only if

- a) $d > 1$ b) $d > 3$ c) $d > 5$ d) $d > 9$

$$\alpha = 0.025$$

$$\text{at } \alpha = 0.025, \chi^2_{critical} = 5$$

H_0 : independent

$$\chi^2_{obs} > \chi^2_{critical} \rightarrow \text{reject } H_0$$

$$\chi^2 = \frac{(2+d)(d)^2}{1 \times (1+d) \times (2) \times 1}$$

$$\chi^2_{obs} = \frac{(d)(2+d)}{2(1+d)} > \chi^2_{critical} = 5$$

$$d > 5 \rightarrow \frac{3(5)(8)}{(2)(7)} = \frac{24}{7} \approx 3.43$$



$$\frac{(10)(12)}{(2)(11)} = \frac{5}{11} \approx 0.45$$

1) Suppose statistics problems and professors come in 3 varieties: uneven, gamy and outgoing. Let's find 76 statistics problems and 63 professors and count how many fall into each variety. This generates the following table:

observed frequencies		
	statistics problems	professors
uneven	35	13
gamy	19	29
outgoing	22	21

Make a table of the expected frequencies.

Using an alpha value of $\alpha=0.05$, test the hypothesis that the statistics problems and professors are distributed independently across the varieties of uneven, gamy and outgoing.

2) Suppose neurons and personality disorders come in 3 varieties: bustling, decorous and tasteless. In your spare time you find 37 neurons and 24 personality disorders and count how many fall into each variety. This generates the following table:

observed frequencies		
	neurons	personality disorders
bustling	8	6
decorous	22	10
tasteless	7	8

Make a bar graph showing the frequencies for all varieties.

Make a table of the expected frequencies.

Using an alpha value of $\alpha=0.05$, test the hypothesis that the neurons and personality disorders are distributed independently across the varieties of bustling, decorous and tasteless.

The time it takes a central processing unit (CPU) to process a certain type of job is normally distributed. A term $\hat{\sigma}_1^2$ is defined as:

$$\hat{\sigma}_1^2 = \frac{1}{15} \sum_{n=1}^{15} x_n^2$$

where x_n represents the observed processing times for each job in a sample of 15 jobs, and the sample mean is 0.

The expected value of $\hat{\sigma}_1^2$ is known to be $\frac{42}{5}$.

What is the probability that the variance of this sample will exceed 12?

$$s^2 = \text{Sample variance} = \frac{1}{14} \sum_{n=1}^{15} (x_n - \bar{x}_n)^2 = \frac{1}{14} \sum_{n=1}^{15} (x_n)^2 = \frac{15}{14} \hat{\sigma}_1^2$$

$$\sigma^2 = E(s^2) = \frac{15}{14} E(\hat{\sigma}_1^2) = \frac{15}{14} \times \frac{42}{5} = \frac{3 \times 6}{2} = 9 // \Rightarrow \sigma = 3$$

① Expectation of Sample Variance = population Variance (σ^2)

$$P(s^2 > 12) = P\left(\frac{s^2(n-1)}{\sigma^2} > \frac{12(n-1)}{\sigma^2}\right) \Rightarrow P(\chi^{n-1} > \frac{12 \times 14}{9}) \rightarrow \text{from table}$$

② if $s^2 = \text{Sample variance}$ then $\frac{s^2(n-1)}{\sigma^2} \sim \chi^{n-1}$

The Chi-Square Statistic for Sample Variance

The formula that transforms a sample variance (s^2) into a variable that follows a chi-square (χ^2) distribution is:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

where:

- n is the sample size.
- s^2 is the sample variance.
- σ^2 is the true population variance (a known or hypothesized value).
- The resulting χ^2 statistic has $\nu = n - 1$ degrees of freedom.

Conditions and Rationale

- Assumption of Normality:** The underlying population from which the sample is drawn must be normally distributed. This is a strict requirement for this relationship to be valid.

- Degrees of Freedom:** The use of $n - 1$ (degrees of freedom) rather than n in the numerator accounts for the fact that the sample mean ($\bar{x}$) is used to calculate the sample variance. Calculating deviations around the sample mean, instead of the true population mean, causes the sample variance to slightly under-estimate the population variance, and the $n - 1$ correction addresses this bias.

$$x_i \sim \text{Uniform}(\mu, \sigma^2)$$

$$E(\hat{\sigma}_1^2) = \frac{1}{15} \sum_{n=1}^{15} E(x_n^2) = \frac{15}{15} (\mu^2 + \sigma^2) = \frac{42}{5}$$

$$\bar{x}_n = \frac{1}{15} \sum_{i=1}^{15} x_n = 0 \rightarrow \text{sample mean}$$

$$\begin{aligned} \frac{x - \mu}{\sigma} &= z \\ x &= \sigma z + \mu \\ x^2 &= \sigma^2 z^2 + \mu^2 + 2\sigma\mu z \\ E(x^2) &= \sigma^2 + \mu^2 \\ \frac{1}{15} \sum_{i=1}^{15} (x_i^2) &= \sigma^2 + \mu^2 \end{aligned}$$

Confidence interval for population standard deviation is

$$\sqrt{\frac{(n-1)s^2}{\chi_{\text{right}}^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_{\text{left}}^2}}$$

$s \rightarrow \text{Sample SD}$

$\chi_{\text{right}}^2 \quad \chi_{\text{left}}^2 \rightarrow \text{critical values from } \chi \text{ table.}$

Spearman's correlation (ρ_s)

→ Convert raw score data of n samples to rank from 1 to n

$$\rho_s = \rho[R[X_i], R[Y_i]] = \frac{\text{Cov}(R[X_i], R[Y_i])}{\sigma_{R[X_i]} \sigma_{R[Y_i]}}$$

Let (1,3), (2,4), (7,8) be three independent observations. Then the sample Spearman rank correlation coefficient based on the above observations is (rounded off to two decimal places).

	$R[X_i]$	$Y[X_i]$
$n_1 \rightarrow 1, 3$	3	3
$n_2 \rightarrow 2, 4$	2	2
$n_3 \rightarrow 7, 8$	1	1

$$\mu_{R[X_i]} = \frac{3+2+1}{3} = 2$$

$$\mu_{R[Y_i]} = 2$$

$$\sigma_{R[X_i]}^2 = \sigma_{R[Y_i]}^2 = \frac{1}{2} (1^2 + 0^2 + 1^2) = 1$$

$$\text{Cov}[R[X_i], R[Y_i]] = \frac{1}{2} ((3-2) \cdot (3-2) + 0 \cdot 0 + (2-1)(2-1)) = 1$$

Shortcut

$$\rho_s = \frac{6 \sum d_i^2}{n(n^2-1)} \rightarrow \text{When all } n \text{ ranks are distinct \& no ties}$$

$$d_i = R[X_i] - R[Y_i]$$

if we use n or $n-1$ according to Bessel's correction, it doesn't matter as the fourth cancels out in calculation as long as we are consistent in numerator & denominator.

$$\rho_s = \frac{\frac{1}{n-1} \sum_{i=1}^n (R[X_i] - \bar{R}_X)(R[Y_i] - \bar{R}_Y)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (R[X_i] - \bar{R}_X)^2} \cdot \sqrt{\frac{1}{n-1} \sum_{i=1}^n (R[Y_i] - \bar{R}_Y)^2}}$$

ρ_s — independent of Bessel's correction applied or not

Pearson Correlation Coefficient $\rho_{X,Y}$

Q) Find $\rho(X,Y)$ if $Y = -3.5X + 12$

Logical short cut $\downarrow$ slope
 $\rho(X,Y) = -1$

$$\sigma_Y = 3.5 \sigma_X$$

$$\text{Var}(Y) = (-3.5)^2 \text{Var}(X)$$

$$E(Y) = (-3.5) E(X) + 12$$

$$E(X)E(Y) = (-3.5) E(X)^2 + 12E(X)$$

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

$$= -3.5 E(X^2) - 3.5 E(X)^2$$

$$= -3.5 \text{Var}(X)$$

$$XY = -3.5X^2 + 12X$$

$$E(XY) = -3.5 E(X^2) + 12 E(X)$$

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

$$\rho(X,Y) = \frac{-3.5 \text{Var}(X)}{3.5 \text{Var}(X)} = -1$$

